BEHAVIORAL AND BRAIN SCIENCES (1989) 12, 321–375
*Printed in the United States of America*

# The child's trigger experience: Degree-0 learnability

**David Lightfoot**
*Department of Linguistics, University of Maryland, College Park, MD 20742*

**Abstract:** According to a "selective" (as opposed to "instructive") model of human language capacity, people come to know more than they experience. The discrepancy between experience and eventual capacity (the "poverty of the stimulus") is bridged by genetically provided information. Hence any hypothesis about the linguistic genotype (or "Universal Grammar," UG) has consequences for what experience is needed and what form people's mature capacities (or "grammars") will take. This *BBS* target article discusses the "trigger experience," that is, the experience that actually affects a child's linguistic development. It is argued that this must be a subset of a child's total linguistic experience and hence that much of what a child hears has no consequence for the form of the eventual grammar. UG filters experience and provides an upper bound on what constitutes the triggering experience. This filtering effect can often be seen in the way linguistic capacity can change between generations. Children only need access to robust structures of minimal ("degree-0") complexity. Everything can be learned from simple, unembedded "domains" (a grammatical concept involved in defining an expression's logical form). Children do not need access to more complex structures.

**Keywords:** cognition, critical periods, development, early experience, evolution, grammar, innate structures, language, learnability, learning, maturation, psycholinguistics

## 1. Introduction: A selective theory of language acquisition

Linguists have traditionally maintained that language is not acquired by children only on the basis of experience; rather, children must themselves contribute something, if only an appropriate "disposition to learn." As careful studies were conducted, so "discovery procedures" and "analogical principles" of increasing complexity were postulated as part of the child's contribution to the process of acquiring a language.

In the last thirty years, generative grammarians have been developing a selective theory of language acquisition. We have sought to ascertain what information must be available to children independently of any experience with language, in order for the eventual mature linguistic capacities to emerge on exposure to some typical "triggering experience." Cutting some corners, we have assumed that this unlearned information is genetically encoded in some fashion and we have adopted (1) as our explanatory model:

(1) a. trigger (genotype → phenotype)
b. primary linguistic data (Universal Grammar → grammar)

The goal is to specify relevant aspects of a child's genotype so that a particular mature state will emerge when a child is exposed to a certain triggering experience, depending on whether the child is raised in, say, a Japanese or Navaho linguistic environment. (1.b) reflects the usual terminology, where "Universal Grammar" (UG) contains those aspects of the genotype directly relevant for language growth, and a "grammar" is taken to be that part of a person's mental make-up which characterizes mature linguistic capacity.

The theory is "selective" in the same sense that current theories of immunology and vision are selective and not "instructive." Under an instructive theory, an outside signal imparts its character to the system that receives it, instructing what is essentially a plastic and modifiable nervous system; under a selective theory, a stimulus may change a system which is already highly structured by identifying and amplifying some component of already available circuitry. Put differently, a selective theory holds that an organism experiences the surrounding environment (and selects relevant stimuli) according to criteria which are already present internally. Jerne (1967) depicts antibody formation as a selective process whereby the antigen selects and amplifies specific antibodies which already exist. Similarly, Hubel and Wiesel (1962) showed that particular neurons were preset to react only to a specific visual stimulus, such as a horizontal line; exposure to a horizontal line entails a radical increase in the number of horizontal line receptors, and a horizontal line can be said to elicit and select specific responses within the organism. Changeux (1980; 1983) argues along similar lines for a theory of "selective stabilization of synapses" whereby "the genetic program directs the proper interaction between main categories of neurons. . . . However, during development within a given category, several contacts with the same specificity may form" and other elements, which are not selected, may atrophy (1980, p. 193). Thus to learn is to amplify certain connections and to eliminate other possibilities (see also Mehler 1974). Jerne (1967) argues that "looking back into

 0140-525X/89 $5.00+.00

the history of biology, it appears that wherever a phenomenon resembles learning, an instructive theory was first proposed to account for the underlying mechanisms. In every case, this was later replaced by a selective theory." For more discussion, see Piattelli-Palmarini (1986) and Jerne's Nobel Prize address (1985). [See also Ebbesson: "Evolution and Ontogeny of Neural Circuits" *BBS* 7(3) 1984.]

Under current formulations of linguistic theory (e.g. Chomsky 1981), the linguistic genotype, UG, consists of principles and parameters that are set by some linguistic environment, just as certain receptors are "set" on exposure to a horizontal line. So the environment may be said to "select" particular values for the parameters of UG. UG must be able to support the acquisition of any human grammar, given the appropriate triggering experience. Of course, UG need not be seen as homogeneous, and may emerge piecemeal, parts of it being available maturationally only at certain stages of a child's development. Grammars must not only be attainable under normal childhood conditions; they must also be usable for such purposes as speech production and comprehension, they must be vulnerable to the kinds of aphasias that one finds, and they should provide part of the basis for understanding the developmental stages that children go through. There is no shortage of empirical constraints on hypotheses about (1).

## 2. Arguments from the poverty of the stimulus

The "logical problem of language acquisition" has provided much of the empirical refinement of (1). Apparent "poverty of the stimulus" problems have led grammarians to postulate particular principles and parameters at the level of UG: The stimulus or trigger experience that children have appears to be too poor to determine all aspects of the mature capacities that they typically attain. It is too poor in three distinct ways: (a) The child's experience is finite but the capacity eventually attained ranges over an infinite domain and must therefore incorporate some recursive property not demanded by experience; (b) the experience consists partly of degenerate data which have no effect on the emerging capacity (see section 4); and, most important, (c) it fails to provide the data needed to induce many principles and generalizations which hold true of the mature capacity. Of these three, (a) and (b) have been discussed much more frequently than (c), although (c) is by far the most significant factor and provides a means for elaborating theories of UG, as I shall now illustrate. For discussion, see Chomsky (1965, ch.1), Hornstein and Lightfoot (1981, pp. 9–31), and Lightfoot (1982, ch.2). [See also Chomsky "Rules and representations" *BBS* 3(1) 1980].

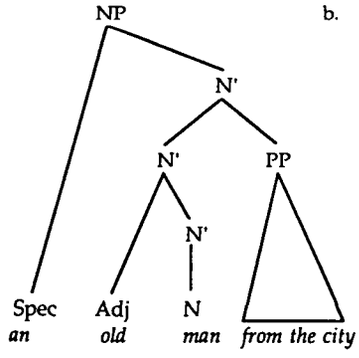
Any argument from the poverty of the stimulus makes crucial assumptions about the nature of the triggering experience. To illustrate, I shall briefly summarize an argument originally presented by Baker (1978), which has been restated in terms of standard X-bar theory by Hornstein and Lightfoot (1981) and by Lightfoot (1982). It has generally been agreed for a long time that linguistic expressions are made up of subunits and have an internal hierarchical structure. It is also generally agreed that a grammar (in the sense defined) is not just a list of expressions but is a finite algebraic system which can "generate" an infinite range of expressions. One might imagine, in that case, that English noun phrases have the structure of either (2.a) or (2.b).

(2) a. NP → Specifier N′
N′ → (Adj) { N′ / N } PP

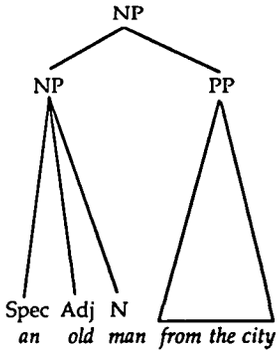
b. NP → NP PP
NP → Specifier (Adj) N

If the phrase structure rules generating a noun phrase (NP) are those of (2.a), a phrase like *an old man from the city* will have the internal structure of (3.a); if the rules are those of (2.b), the structure will be (3.b). In (3.a) *an old man*, for example, does not form a single phrasal unit, but in (3.b) it does. The crucial difference is that the rules of (2.a) refer to N′, an element intermediate between the head noun (N) and the maximal phrasal projection (NP) of that noun.

(3) a.




b.




Now, it can be shown that any noun phrase that occurs in English, and thus any noun phrase that an English-speaking child will hear, can be generated by both sets of rules. However, linguists believe that something along the lines of (2.a) must be correct, or at least preferred to (2.b), because (2.b) is consistent with certain phenomena which do *not* occur in English (in contrast to (2.a)) and because (2.b) has no N′ node, and therefore provides no straightforward way to distinguish between (4.a&b), and no ready means to capture the ambiguity of (5.a), which may have the meaning of (5.b) or (5.c). The details of the analysis need not concern us here.[1]

(4) a. *the student of physics is older than the one of chemistry
b. the student from New York is older than the one from Los Angeles

(5) a. he wants an old suit but he already has the only one I own
b. he wants an old suit but he already has the only suit I own
c. he wants an old suit but he already has the only old suit I own

What *is* relevant here is the following problem: Children might be exposed to any noun phrase that could occur in English, but they are not systematically informed that sentences like (4.a) are not uttered by most speakers and that (5.a) has two possible meanings. In fact, perception of ambiguity is a sophisticated skill which develops late and not uniformly; most ambiguities pass unnoticed and people take the most appropriate of the available meanings. Children do come to know these things, and this knowledge is indeed part of the output of the language acquisition process, but it is not part of the

input, not part of the "evidence" for the emerging system, and thus not part of the triggering experience. Consequently, although linguists are able to choose hypothesis (2.a) over (2.b) on the basis of phenomena like (4) and (5), children have no analogous basis for such a choice if such data are not available to them. It is in this sense that the stimulus is too impoverished to fully determine the emergent analysis.

Hence children must arrive at (2.a) on some other, presumably nonexperiential basis. As an account of this basis, linguists have postulated genotypical information that phrasal categories have the structure of (6).

(6) a. XP → Specifier, X′
b. X′ → $\begin{Bmatrix} X' \\ X \end{Bmatrix}$, (YP)

According to (6.a), any noun phrase (NP) consists of a Specifier (e.g. an article) and a subphrase N′ in some order to be determined by the child's particular linguistic experience, the "trigger" of (1.a); similarly, a verb phrase (VP) consists of a Specifier and a V′ in some order, and likewise the other phrasal categories. By (6.b), the N′ consists of a head (N or N′) and perhaps some complement phrase in some order (the comma indicates an unordered set and YP covers any phrasal category or a clause).

Under (6), the linear order of constituents constitutes a parameter that is set on exposure to some trigger, for example (7).

(7). a. the house
b. students of linguistics, belief that Susan left

The English-speaking child hears phrases like (7.a) and, after some development, analyzes them as consisting of two words, one of a closed class (*the*) and the other of an open class (*house*); in the light of this and in the light of the parameter in (6.a), the child adopts the first rule of (2.a). Likewise, exposure to phrases like (7.b) suffices to set the parameter in (6.b), so that the second rule of (2.a) is adopted.[2] Given the parameters in (6), rules like those of (2.b) are never available to children and therefore do not have to be "unlearned" in any sense. Although no "evidence" for the existence of a phrasal category N′ seems to be available in a child's experience, it is provided by the genotype and therefore occurs in mature grammars (I shall consider an alternative account later).

There is much more to be said about this argument and about its consequences. I have sketched it briefly here in order to demonstrate that a poverty-of-stimulus argument is based on assumptions about the triggering experience. The assumption so far has been that the nonoccurrence for many people of (4.a) and the ambiguity of (5.a) are not part of the trigger, but that garden variety NPs like (7) are. It should be clear that there is a close relationship between the three entities of (1), and a claim made about any one of them usually has consequences for hypotheses about the other two.

This is by now reasonably well-established for UG and particular grammars, but there has been a curious silence about the triggering experience. Generativists nowadays describe "parametric differences" between the grammars of, say, Japanese and Navaho, but they rarely mention how the parameters would be set for the particular grammars of these languages: what the triggering experience would need to be for the Japanese and Navaho child. Worse, if one tries to tease out the implicit assumptions about the trigger, they sometimes include exotic or negative data (see below). Despite this incomprehensible omission, which potentially undermines the claims being made, there is an intrinsic relationship between the items in (1). If the trigger or the "primary linguistic data" (PLD) were rich and well-organized, correspondingly less information would be needed in UG, and vice versa. These are not aesthetic swings and roundabouts; there are clear facts which limit viable hypotheses. My goal here is to focus attention on the triggering experience.

I shall argue that the trigger consists of a haphazard set of utterances made in an appropriate context, utterances of a type that any child hears frequently. In other words, it consists of robust data and includes no "negative data"– no information that certain expressions do *not* occur. First I will contrast this idea with some other ideas in the literature. Then in section 5 I will flesh it out, making it more precise and arguing that the trigger consists only of simple, unembedded material and that everything can be learned from structures of "degree-0" complexity, where structural complexity is defined in terms of logical forms.

## 3. Negative data

It is clear that the PLD which trigger the growth of a child's grammar do not include much of what linguists use to choose between hypotheses. To this extent, the children are not "little linguists," constructing their grammars in the way that linguists construct their hypotheses. For example, the PLD do not include well-organized paradigms or comparable data from other languages. Nor do the PLD include rich information about what does not occur, that is, negative data.[3] It is true that some zealous parents correct certain aspects of their child's speech and so provide negative data, but this is not the general basis for language development. First, such correction is not provided to all children and there is no reason to suppose that it is an indispensable ingredient for language growth to take place. Second, even when it is provided, it is typically resisted, as many parents will readily attest. McNeill (1966, p.69) recorded a celebrated illustration of this resistance.

Child: Nobody don't like me.
Mother: No, say "nobody likes me."
Child: Nobody don't like me.
(eight repetitions of this dialogue)
Mother: No, now listen carefully; say "nobody likes me."
Child: Oh, nobody don't likes me.

Third, correction is provided only for a narrow range of errors, usually relating to morphological forms. So, the occasional *taked, goed, the man what we saw,* etc, might be corrected; on the eighth try McNeill's child perceived only a morphological correction, changing *like* to *likes*. However, not even the most conscientious parents correct deviant uses of the contracted form of verbs like *is* and *will* (8) – in this case because they do not occur in children's speech.

(8) a. *Jay's taller than Kay's (cf. . . . than Kay is)
b. *Jay'll be happier than Kay'll (cf. . . . than Kay will)

Parents also do not correct errors in which anaphors such as *each other* are misused. Matthei (1981) reports that children sometimes interpret sentences like *the pigs said the chickens tickled each other* with *each other* referring to *the pigs*. This misinterpretation, discovered under experimental conditions, is unlikely to be perceived by many adults in everyday circumstances. Similarly with many other features of children's language; for good discussion, see Baker (1979).

It is sometimes argued that although children are not supplied with negative data directly, they may have access to them indirectly. So Chomsky speculates along these lines:

> If certain structures or rules fail to be exemplified in relatively simple expressions, *where they would be expected to be found* [my emphasis], then a (possibly marked) option is selected excluding them in the grammar, so that a kind of "negative evidence" can be available even without corrections, adverse reactions, etc. (1981, p.9)

This is illustrated by the so-called null-subject parameter, whereby finite declarative sentences like (9) occur with a phonetically null subject in Italian, Spanish, and many other languages, but the corresponding expressions do not occur in English, French, and so on.

(9) a. ho trovato il libro
b. chi credi che partirà?
c. *found the book
d. *who do you think that will leave?

Whatever the form of the parameter permitting this kind of variation, Chomsky, following Rizzi (1982b), suggests that if the English-speaking child picks the wrong setting, then failure to hear sentences like (9.c) might be taken as indirect evidence that such sentences are ungrammatical and thus do not occur for some principled reason. Consequently the child will pick the setting which bars (9.c,d).

Two remarks on this: First, if children do have indirect access to negative data, it will have to be specified under what circumstances this occurs. That is, in Chomsky's formulation above, the phrase "in relatively simple expressions, where they would be expected to be found" will need to be elaborated in such a way that it distinguishes cases like (9) from those like (4) and (8). One can argue that children may have indirect access to data like (9.c), but it is hardly plausible to say that they have indirect access to (8). For this distinction to be made, UG would have to be enriched to include analogical notions which have not yet been hinted at.

Second, so far as there are no strong arguments for indirect access to negative data. There are certainly plausible alternative explanations for the null-subject phenomenon. One possibility is to claim that the English setting for the relevant parameter is *unmarked*, that is, the default case. Thus Italian and Spanish children need specific evidence to adopt the other setting, and (9.a) is the required evidence.[4] The fact that the Italian setting for the parameter seems to be much more common across languages than the English setting does not entail that it is less marked, since markedness values do not reflect statistical frequency. In fact, Berwick's Subset Principle (1985) predicts that the Italian setting should be marked. The Subset Principle requires children to "pick the narrowest possible language consistent with evidence seen so far" (p.237). The Italian setting of the parameter entails a language which is broader than one with the English setting (because in Italian subjects may or may not be phonetically expressed), and therefore the English setting needs to be unmarked (p.290).

Another possibility is to make the variation in null subjects depend on some other property. It has often been suggested that null subjects occur only in grammars with rich verbal inflection. However, rich inflection is not a sufficient condition for null subjects: German does not have null subjects, although its verbal inflection involves number, person, and gender like that of Spanish, which does allow null subjects. Consequently, the learning problem remains constant and is unaffected by the richness of inflections. As an alternative account, Hyams (1983) related the impossibility of null subjects to the occurrence of expletive pronouns (*it's cold, there's no more*), which occur in English, French, and German but not in Italian or Spanish; she marshalled some interesting evidence in favor of something along these lines by considering the developmental stages that children go through.

Indirect access to negative data may prove to be necessary for a full explanation of language acquisition but so far no very plausible case has been made.[5] The notion raises nontrivial problems in defining the contexts in which indirect access is available. Meanwhile, plausible solutions for problems which seem to call for indirect access to negative data may be suggested by viewing the phenomena in relation to other properties and not in isolation. I have mentioned two such possibilities here, but there are other suggestions in the literature.

## 4. Not all experience is a trigger

Putting aside further discussion of the possibility of indirect access to certain negative data, one can plausibly argue that the triggering experience is less than what a "little linguist" might encounter and that it does not include information about ungrammatical sentences or the many other properties that are analyzed in a typical issue of *Linguistic Inquiry*. Such information is simply not part of any child's linguistic experience. Consequently, we may persist with the idea that the trigger consists of nothing more than a haphazard set of utterances in an appropriate context. In fact, we can restrict things further: The trigger is something less than the total linguistic experience. Neither the occasional degenerate data that a child hears nor idiosyncratic forms necessarily trigger some device in the emergent grammar which has the effect of generating those forms. So, for example, a form like (10.a) might occur in a child's experience without triggering the formation of a rule that generates this unusual form of subject–verb agreement. Similarly, a child growing up in New York might hear (10.b) without adopting *y'all* as a word in his grammar.

(10) a. the person who runs the stores never treat people well
b. y'all have a good time in South Carolina

A child might even be exposed to significant quantities of linguistic material that does not act as a trigger. So, if a house-guest speaks an unusual form of English, perhaps with different regional forms or the forms of somebody

who has learned English imperfectly as a second language, this normally has no noticeable effect on a child's linguistic development. Even children of immigrant parents with heavy accents perpetuate few nonstandard aspects of their parents' speech.

This is fairly obvious intuitively and shows that there is little to be learned about the trigger experience from simply tape-recording everything uttered within a child's hearing (cf. Wells 1981). More can be learned from the historical changes that languages undergo. It is well known that certain kinds of syntactic patterns become obsolete in certain speech communities at certain times. This means that children sometimes hear a form which does not trigger some grammatical device for incorporating this form in their grammar. Thus, even though they have been exposed to the form, it does not occur in their mature speech. The conditions under which this happens cast some light on the nature of the trigger.

Consider an example discussed in some detail in Lightfoot (1979; 1982). In earlier forms of English, sentences such as *the king likes the queen* used to occur with the same general meaning that they have today but with *the king* being analyzed as the object of the verb *like* and *the queen* as the subject. So a Middle English speaker might also have heard (11), where the postverbal noun phrase is clearly the subject (having an appropriate case in (11.a) and agreeing with the verb in number in (11.b)). It is plausible to assign such forms structure (12), where the postverbal subject is linked to the canonical subject position, which is empty (and perhaps the residue of a movement transformation). *Like* must have meant "cause pleasure for," unlike in Modern English, where it always means "derive pleasure from." There were some forty or fifty verbs which could occur in such a syntactic context: Other examples were *repent, rue, ail,* and so forth.

(11) a. him likes the queen
b. the king like the pears

(12) $_{NP}[e_i]\ _{VP}[\ _{V'}[\text{him likes}]\ _{NP}[\text{the queen}_i]]$

Forms like (11) died out during the Middle English period; that is, some speakers must have heard them but did not reproduce them. For those speakers such forms were part of their linguistic environment but were not part of the trigger experience; they did not trigger a grammatical device which permitted their generation, and we know why. Notice that structure (12) contains a V′ with object–verb order. This order was characteristic of early English, just as it is characteristic of modern Dutch and German. Given the phrase structure convention of (6.b), speakers of early English had the head of V′ in final position, whereas later English speakers fixed the order differently. For a discussion of how that change took place, see Canale (1978) and Lightfoot (1979; 1982). My point here is that as the order of elements within the V′ changed, so forms like (11) were no longer part of the trigger, although they were heard by children. When the order of constituents in V′ was fixed differently, a form like (11.a) would have a structure like (13).

(13) $_{NP}[e_i]\ _{VP}[_{V'}[\text{likes him}]\ _{NP}[\text{the queen}_i]]$

The problem is that there is no ready way for *him* to move to a preverbal position without violating a principle of grammar. There is only one available NP position for *him* to move to, since there is no evidence that Middle English allowed a preverbal clitic NP. If *him* moved to the empty subject position, the indexed empty element would no longer exist, and there would be no way of interpreting *the queen* as the subject with which the verb must agree in number and person. The derivation would violate Chomsky's (1981) "theta criterion", because there would be no one-to-one relationship between noun phrases and "theta positions," that is, "logical" positions in which an NP may be understood as filling a semantic function (*him* would be associated with two positions and *the queen* with none). Consequently, a form like (11.a) could not be derived from a structure like (13).

Nonetheless, children at the relevant stage presumably heard sentences like *the king likes the queen* uttered in a context in which it was clear that the king was happy and that the queen was the reason for his happiness. Since an analysis like (13) was not available, a ready alternative was adopted where no NP was moved and the preverbal NP was interpreted as the subject (14). Forms like (11) were heard but not reproduced; they were replaced by (14.b,c).

(14) a. $_{NP}[\text{the king}]\ _{VP}[\ _{V'}[\text{likes}\ _{NP}[\text{the queen}]]]$
b. $_{NP}[\text{he}]\ _{VP}[\ _{V'}[\text{likes}\ _{NP}[\text{the queen}]]]$
c. $_{NP}[\text{he}]\ _{VP}[\ _{V'}[\text{likes}\ _{NP}[\text{the pears}]]]$

Under this syntactic analysis, which was forced for the reasons given, and given the perceived meaning of the sentence, *like* could only be interpreted with a different meaning, the modern "derive pleasure from." Consequently, the old meaning of *like* and forms like (11) were not part of the trigger experience and died out of the language; the reason was that they could not be interpreted consistently with the context in which they were uttered and consistently with a V′ containing a head preceding a complement . . . assuming that other aspects of the grammar were triggered as in the immediately preceding generation of speakers.[6]

The point of this discussion has been to reinforce the claim that the trigger experience may be less than the total linguistic experience of a child, and to show that considering the conditions under which structures become obsolete in a language tells us something about the limits to trigger experiences. Indeed, from the point of view of the research program under discussion here, this seems to be a major interest of historical change: Properly interpreted, it can reveal something about the limits to trigger experiences if one studies the conditions under which structures drop out of the language and thus fail to be triggered at a certain stage of the history of a language.

The trigger experience is hence some subset of a child's total linguistic experience. But where exactly are the limits? This is often a crucial question in grammatical analyses, but it is rare to see alternatives discussed. Consider again the example of the structure of noun phrases. I argued above that any noun phrase that an English-speaking child could hear would be consistent with the rules of both (2.a) and (2.b). I also claimed that the data which lead grammarians to prefer (2.a) to (2.b) are generally not available to children and that the information which eliminates (2.b) must therefore come from some other, presumably genetic source. One could look at things somewhat differently, however. The real difference between (2.a) and (2.b) is the existence of a N′ node in the rules of (2.a). The existence of this node is

required by the UG rule schema in (6) and, on that account, does not have to be derived from relevant experience. In that case, we might ask whether there is anything in a child's experience which would require postulating an N′ node; one can indeed imagine evidence that would force the child to establish such a node.

English speakers use the indefinite pronoun *one* to refer back to an N′ (see note 1); the fact that it refers to an N′, something intermediate between the head N and its maximal projection (NP), might in fact be learnable. A sentence like (15.a) would not be a sufficient basis for learning this because, regardless of whether Heidi actually has a big or small cup, the sentence could always be interpreted as specifying only that Heidi had some cup regardless of size (with *one* referring only to the N *cup*). Sentence (15.b), however, would suffice if uttered in a situation where Heidi has a cup that is some color other than blue; only the interpretation with *one* representing *blue cup* would be consistent with the facts. In that case, a child might *learn* correctly that *one* stands for a phrasal projection of N, namely an N′.[7]

(15) a. Kirsten has a big cup, and Heidi has one too
b. Kirsten has a blue cup, but Heidi doesn't have one

We now have two alternative accounts: The existence of N′ might be derived from a property of UG or it might be triggered by the scenario just sketched. My hunch was and remains that this scenario is too exotic and contrived to be part of every child's experience, and therefore that postulating (6) at the level of UG is more plausible. But this hunch may be wrong. It is certainly falsifiable. If a rule schema like (6) exists in UG, then strong claims are made about the possible degree of variability in the structure of NPs that will be found in the languages of the world: In languages where this kind of structural configuration is relevant (which may or may not be *all* languages), there will be essentially four NP types as in (16).

(16) a. $_{NP}$[Spec $_{N'}$ [head complement]]
b. $_{NP}$[Spec $_{N'}$[complement head]]
c. $_{NP}$[ $_{N'}$[head complement] Spec]
d. $_{NP}$[ $_{N'}$ [complement head] Spec]

Type (16.a) is represented by English, French, and so on, and type (16.b) seems to be manifested in Basque, Burmese, Burushaski, Chibcha, Japanese, Kannada, and Turkish (see Greenberg 1966, n. 20). Types (16.c,d) are more problematic because I know of no carefully studied grammar which manifests them. Greenberg (1966) and Hawkins (1979) discuss several languages in which demonstratives follow the head noun and which therefore might be of type (16.c or d), but they do not distinguish between demonstratives which have the syntax of adjectives (as in Latin) and those which manifest Spec (as in English). If it should turn out that types (16.c,d) do not occur, then the rule schema of (6.a) will be tightened to allow only the Spec-N′ order.

Also, the rule schema (6) suggests that one will find developmental stages corresponding to the fixing of the two parameters by a child. Lightfoot (1982, p.179f), building on work by Klima and Bellugi (1966) and Roeper (1979), argues that this is indeed the case. Children seem to acquire noun phrase structures in four identifiable stages. Examples (17.a,b) list some noun phrases occurring in the first two stages.

| | | | | | |
|---|---|---|---|---|---|
| (17) | a. | car | b. | a coat | that Adam |
| | | baby | | a celery | more coffee |
| | | wa-wa (water) | | a Becky | two socks |
| | | mama | | a hands | big foot |
| | | hands | | my mommy | |

All children go through the four stages at some point, although the ages may vary. Most children utter the stage 2 forms between one and two years. At stage 3 there is more sophistication.

| | | | |
|---|---|---|---|
| (18) | mama | my doll | a blue flower |
| | cracker | your cracker | a nice cap |
| | doll | | a your horse |
| | spoon | | that a horse |
| | | | that a blue horse |
| | | | your blue cap |

At stage 4 the mature system emerges, which normally remains more or less constant for the rest of the child's lifetime. But consider (19), some forms that never occur in children's speech.

| | | | |
|---|---|---|---|
| (19) | *blue a flower | *a that blue flower | *flower a |
| | *nice a cup | *blue a that | *house that a |
| | | *my a pencil | *that a |
| | *a that house | *a my | *my a |

Recall the parameters for noun phrases developed earlier. These were hypotheses about how NP structure could vary from grammar to grammar. At stage 1 these parameters are irrelevant, because the child has only one-word structures. Other cognitive capacities are relevant, such as the conceptual system that involves properties and conditions of reference, knowledge and belief about the world, conditions of appropriate use, and so on. These play a role in explaining why *mama* and *cup* are more likely than *photosynthesis*, *quark*, or *grammar* to be among the earliest words in a child's speech.

At stage 2 the child seems to have fixed the linear order parameter of rule schema (6.a) and determined that the order is Spec N′: All specifiers appear at the front of the noun phrase. The occurrence of phrases like *a Becky* or *a hands* suggests that at this stage children cannot distinguish definite and indefinite articles, and that they do not know that *a* is singular. There is no evidence that the child can distinguish subtypes of specifiers (articles, possessives, numerals, demonstratives), but they all occur one at a time in front of a noun.

By stage 3, children discriminate some kinds of specifiers and establish some more of the relative orders. In fact, the child knows that all specifiers precede adjectives, which in turn precede nouns, and that specifiers are optional, whereas the noun is obligatory. The stage 3 grammar differs from the mature system in that the child does not yet know that an article may not co-occur with a demonstrative or with a possessive like *your*. This suggests that the child now has the rules NP → Spec N′, N′ → (Adj) N, but that it takes a little longer to determine the status of a demonstrative and whether a form like *your* is a specifier or an adjective. After all, in other languages demonstratives and possessives are often adjectives instead of specifiers.

Consequently, there is reason to believe that postulating the rule schema in (6) at the level of UG is more plausible than claiming that the existence of N′ is *learned* on the basis of exposure to sentences like (15.b) uttered in

the relevant context. But the important point is that alternatives like this need to be sketched and evaluated, and that grammarians should be paying more attention to their assumptions about the nature of the triggering experience required to set the parameters of UG that they hypothesize.

So I persist with the idea that the trigger is a subset of a child's experience, and that it probably does not include exotic events like the one sketched above in the context of (15.b). The trigger consists only of robust data which can be analyzed consistently with genotypical principles. There remains the question of how small the subset is.

There is a theory, advanced recently by Snow (1977) and earlier by Jespersen (1922, p.142; as pointed out to me by Jim McCawley), that the crucial input for language growth to take place is very small: a specially structured form of speech transmitted through mothers and caretakers. This "motherese" is supposed to provide a set of patterns which are generalized by children on an inductive basis. This view was held fairly widely for a while.

There are at least four reasons why this kind of pattern generalization is not the answer to how children acquire speech. First, although children no doubt register only part of their linguistic environment, there is no way of knowing quite what any individual child registers. Hence factual basis is lacking for the claim that children register only what is filtered for them through parents' deliberately simplified speech. Children have access to more than this, including defective utterances.

Second, even supposing that children register only perfectly well-formed expressions, this would not be enough to show that the child has a sufficient inductive base for language acquisition. Recall that the child's stimulus is "deficient" in three distinct ways (sect. 2 above); the motherese hypothesis would circumvent only the degeneracy problem (b) but leaves untouched the infinity problem (a) and the far more important problem (c): the absence of evidence in PLD (primary linguistic data) for certain partial generalizations. The poverty of stimulus problems still hold and the child would need to know that the contractability of the first *is* in (8) could not be extended to the second *is*. One wants to know why quite ordinary inductive generalizations like this are not in fact made; the so-called motherese does not show where inductive generalizations must stop.

Third, if the child registered only the simplified and well-formed sentences of motherese, the problem of language learning would be *more* difficult because the child's information would be more limited.

Fourth, careful studies of parents' speech to children (e.g. Newport et al. 1977) show that an unusually high proportion consists of questions and imperatives; simple declarative sentences are much rarer than in ordinary speech. This suggests that there is very little correlation between the way the child's language emerges and what parents do in their speech directed at children. Thus, the existence of motherese in no way eliminates the need for postulating a genetic basis to explain language acquisition. The child is primarily responsible for the acquisition process, not parents or older playmates. (For a good discussion of this topic, see Wexler and Culicover 1980, pp. 66–78.)

Furthermore, though it is by no means clear exactly what this motherese consists of, the general phenomenon is not uniform and does not occur in all households or cultures. Even where motherese is not practiced, children nonetheless attain a normal linguistic capacity. This suggests that the child's trigger experience does not need to be limited artificially along the lines of motherese.

## 5. Degree-0 learnability

Wexler and Culicover (1980) argued for limits to a child's trigger with their notion of degree-2 learnability. They offered the first "learnability proof," showing that, given certain constraints on grammatical processes, a child might be confronted with (*b*, *s*) pairs (where *b* is a base structure, in which transformational and phonological rules have not applied, and *s* a surface string) and would identify the relevant grammar in finite time when exposed only to sentences with two levels of embedding; this is "degree-2 learnability." The first learnability proof was bound to be important and many interesting assumptions were made about language learners, not all of them very plausible (see Baker 1982; Morgan 1986). In particular, the notion that children must have access to very complex sentences seems worth examining.[8]

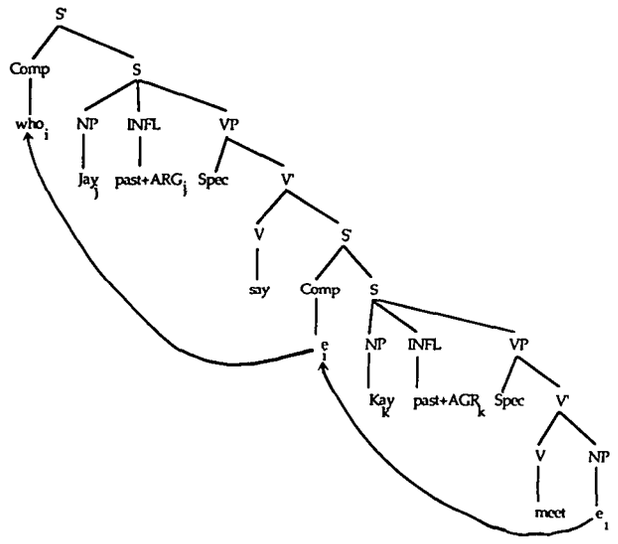
Since Wexler and Culicover there have been real advances in grammatical theory, notably the development of the parameter-setting model of UG, replacing the earlier selectional model of Chomsky (1965), whereby a child was said to "select" the appropriate grammar through the aid of an evaluation metric. Also, for the last fifteen years grammarians have been trying to develop locality restrictions in such a way that grammatical processes affect only elements which are not too far apart. This work suggests that, in general, grammatical processes have effects only on items which are clause-mates (members of the same clause) or where an item in a lower clause is, loosely, at the front of that clause. Locality restrictions are formulated somewhat differently at different stages of the development of UG and by different authors. The details of various locality restrictions need not concern us immediately, but they do raise the following question: If grammatical processes are generally limited to clause-mates or at most to items of which one is at the front of an embedded clause, why should children need to hear more than a single clause (plus the front of a lower clause) in order to hear the effects of all possible grammatical processes in their language? In other words, can everything be learned from main clauses (degree-0 learnability) plus a little bit?

This work may *suggest* a tight restriction on the trigger experience, but there is no demonstration effect (although this has sometimes been assumed, e.g. Culicover & Wilkins 1984; see Lightfoot 1986 for discussion). This will be illustrated as we consider some apparent counterexamples to degree-0 and even degree-1 learnability, which are consistent with the usual locality conditions and with degree-2 learnability. I shall argue that these cases are in fact compatible with a degree-0 learnability account.

**5.1. Bounding nodes.** Chomsky (1973) proposed a Subjacency condition, which requires syntactic movement to be local, crossing no more than one "bounding" node. Wh-phrases move to a complementizer position

("Comp") at the front of a clause; hence a clause, S′, consists of Comp and S, and S in turn consists of a subject NP, an inflection element INFL (including tense and an agreement marker), and a VP. Consequently, a wh-phrase must move step-wise from within an embedded clause to the front of a main clause, leaving a "trace" at each stage; so *who did Jay say Kay met?* would be derived as in (20), and each movement crosses only one bounding node, S.

(20)




Rizzi (1982a, p.73 n.25) and Sportiche (1981) have argued for a parametric difference between English on the one hand and French and Italian on the other, so that English speakers have S as a bounding node whereas French and Italian speakers have S′. Rizzi points to (21.a) and Sportiche to (21.b,c), which have a long-distance movement as in (22) (which is Sportiche's analysis for 21.c).

(21) a. tuo fratello, a cui mi domando che storie abbiano raccontato, era molto preoccupato
"your brother, to whom I wonder which stories they told, was very troubled"
b. c'est à mon cousin que je sais lequel offrir
"it's to my cousin that I know which one to offer"
c. voilà une liste des gens à qui on n'a pas encore trouvé quoi envoyer
"here's a list of the people to whom we've not yet found what to send"

(22) [à $qui_i$] $_S$[on n'a pas encore trouvé $_{S'}$[[$quoi_j$] $_S$[envoyer $e_j$ $e_i$]]]

In (22) *à qui* moves over two S nodes but only one S′ node. Consequently, if the Subjacency condition proscribes movement over more than one bounding node, (21) suggests that in French and Italian S′ and not S must be the bounding node. If we assume that the English equivalents of (21) are ungrammatical (see Rizzi 1982a and Sportiche 1981), then English grammars have S as a bounding node and a movement like that of (22) would violate Subjacency. Rizzi (1982a) is one of the rare instances of a grammarian discussing the needed trigger experience; he argues that if the Subjacency condition is parameterized so that either S or S′ or both may be a bounding node, then S must be a less marked value than S′. The reason is that if S′ is unmarked, then it is hard to see how the English-speaking child could learn the S value without having access to negative data, namely, that the English analogues to (21) do not occur. On the other hand, the Italian and French child could learn the S′ value by hearing sentences like (21). Hence a trigger is provided, but the trigger is a complex structure with two levels of embedding, which is consistent with degree-2 learnability but not with degree-0.[9]

Continuing to assume Rizzi and Sportiche's analysis and its factual basis, one may ask whether there is an alternative way to learn what is needed: that S′ and not S is a bounding node in French and Italian. A simple sentence like (23.a) might appear to provide the relevant trigger if it is analyzed along the lines of (23.b). In (23.b) there is movement over NP and S, suggesting that both cannot be bounding nodes. However, there is an alternative analysis in which movement takes place from outside the NP, as in (23.c), as Hornstein and Weinberg (1981) showed was needed for English (23.d). If (23.c) is a possible analysis, then (23.a) would not provide crucial evidence (i.e., a trigger) for the choice of S′ as a bounding node as opposed to S.

(23) a. de qui as-tu vu un photo?
b. de $qui_i$ $_S$[as-tu vu $_{NP}$[un photo $e_i$]]
c. de $qui_i$ $_S$[as-tu vu $_{NP}$[un photo] $e_i$]
d. $who_i$ $_S$[did you see $_{NP}$[a photo] of $e_i$]

However, there are simple main clause data which require that S not be a bounding node; (24.a) needs to be analyzed as (24.b), as argued by Kayne (1981) and others. This movement of *combien* from within an object NP, across both NP and S, dictates that S cannot be a bounding node in addition to NP.[10]

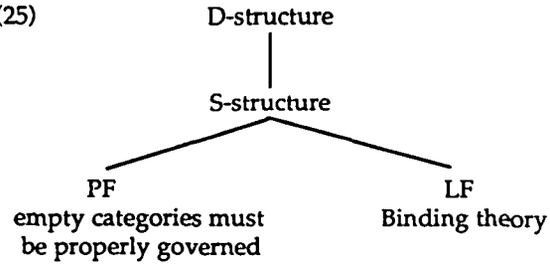
(24) a. combien as-tu vu de personnes?
"how many people have you seen?"
b. $combien_i$ $_S$[as-tu vu $_{NP}$[$e_i$ $_{N'}$[de personnes]]

I have argued here that at least for French (I shall return to Italian in section 5.4 below) the relevant parametric value can be triggered by unembedded data like (24), and therefore that Rizzi and Sportiche's analysis of (21) does not motivate degree-2 or degree-1 learnability. However, it is important to note that whether or not there was a degree-0 trigger for the bounding node parameter is quite unrelated to the existence of locality conditions, which do not ipso facto require degree-0 learnability.

I shall consider two further examples of analyses which seem to require a more complex trigger than would be permitted under degree-0 learnability, but I shall argue that the relevant parameters may be set on the basis of unembedded material. These examples require a discussion of some theoretical machinery.

For both cases I shall adopt the usual model of grammar (25) and the specific UG proposals of Weinberg et al. (1987). Under (25), a D-structure (which represents the underlying grammatical relationships) is mapped onto an S-structure through successive movement operations. S-structure, in turn, is mapped onto both a phonological form (PF), which is the basis for a phonetic representation, and a logical form (LF), which represents a significant part of the meaning of an expression, indicating the scope relations of quantifiers, anaphoric relations among NPs, and so forth.

(25)




Weinberg et al. propose two locality conditions. They argue that an anaphor must be *bound* (i.e. coindexed with some higher element) within its Domain at logical form (26.a), and that this is part of a more general binding theory along the lines of Chomsky (1981).

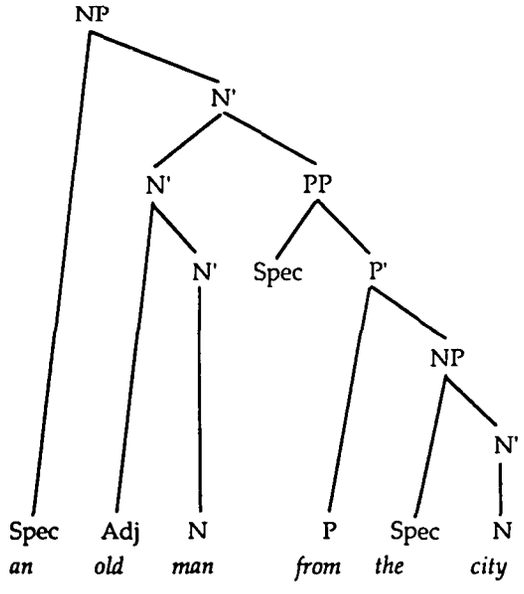
(26) Binding theory
- a. an anaphor is bound in its Domain
- b. a pronoun is not bound in its Domain
- c. a referential expression is not bound by an argument (for our purposes, a NP inside S)

They also require an empty element in PF to occur only where it is *governed* by a coindexed element. A government relation holds if two elements share all maximal projections; more technically, (27).

(27) α (X⁰ or head in our earlier terminology) governs β
i. if all maximal projections dominating α also dominate β *and* ii. α is dominated either by
a. all maximal projections dominating β
or b. by all maximal projections dominating the maximal projection of β.

Consider how (27) applies with respect to a more detailed version of (3.a), given as (3′) below.

(3′)




NP and PP are maximal phrasal projections of the heads N and P respectively. In (3′), the relationship between the N *man* and Spec *an*, Adj, and PP satisfies the conditions of (27i and ii.a). Thus the N *man* governs these elements. It also governs the P *from* via (27i and ii.b) since its maximal phrasal projection PP is dominated by all the maximal phrasal projections dominating the N *man*. However the N *man* does not govern the N *city* because its maximal phrasal projection NP is dominated by another maximal phrasal projection PP which does not dominate the N *man*, in violation of (27ii.b).

**5.2. Dutch government.** This formulation of UG permits a degree-0 account of some phenomena which have been widely discussed for the last ten years. As often noted, English speakers do not move subjects from a position immediately to the right of an overt complementizer. Structures like (28.b) violate the UG condition that empty elements at PF must be governed by a coindexed element. Hence, corresponding sentences like **who did Jay say that saw Kay?* do not occur in English (by contrast, sentences corresponding to (28.a) and (28.c) do occur: *who did Jay say that Kay saw?* and *who did Jay say saw Kay?*).

(28) a. $_{S'}$[who$_i$ did+AGR$_j$ Jay$_j$ say $_{S'}$[ $_{Comp}$[e$_i$ that] Kay$_k$ AGR$_k$ saw e$_i$ ]]
b. *$_{S'}$[who$_i$ did+AGR$_j$ Jay$_j$ say $_{S'}$[ $_{Comp}$[e$_i$ that] e$_i$ AGR$_i$ saw Kay ]]
c. $_{S'}$[who$_i$ did+AGR$_j$ Jay$_j$ say $_{S'}$[ $_{Comp}$[e$_i$ ] e$_i$ AGR$_i$ saw Kay]]

Dutch speakers, however, use sentences equivalent to (28.b): *wie denk je dat het boek gelezen had* "who do you think that had read the book." Such sentences have a structure like (29.a): The Subjacency condition forces *wie* to move first to its local Comp leaving a trace there which makes the empty subject governed via a Comp indexing procedure.[11] Similarly, in structures like (29.b) the indexed Comp properly governs an empty subject despite the presence of the complementizer.

(29) a. wie$_i$ denk je $_{S'}$ [$_{Comp}$[e$_i$ dat], $_S$[e$_i$ het boek gelezen had] ]
who think you that the book read had
b. ik vraag me af $_{S'}$ [$_{Comp}$[wie$_i$ ofdat], $_S$[e$_i$ het boek gelezen had] ]
I wonder who whether the book read had
"I wonder who read the book"

Koopman (1983) discussed these phenomena and raised the question of what the relevant parameter was that distinguished Dutch and English grammars; she further asked how that parameter could be set, arguing that it could be set on the basis of main clause data. The essential idea is that a Dutch speaker has to learn that an item in Comp may percolate its index to Comp and make an empty subject governed despite intervening material, unlike in English, where percolation is blocked by the presence of other material in Comp; so in (28.b) the trace in Comp may not percolate its index and hence the subject trace is not properly governed. Koopman argued that this distinction can be learned from main clause data like (30). Dutch has an underlying object–verb order and *heeft* must therefore move toward the front of the sentence, as indicated. The empty subject must be governed by the Comp despite the presence of *heeft*, just as the empty subjects in (29) are governed by the coindexed Comp despite the *dat* and *ofdat*.

(30) $_{Comp}$[wie$_i$ heeft$_j$]$_i$ $_S$[e$_i$ het boek gelezen e$_j$]
who has the book read
"who has read the book?"

English, however, does not have structures like (30), and thus no main clause evidence that an item in Comp may effectively govern across intervening material (via the Comp indexing procedure). So sentences like *who has read the book* have a structure like (31.a), where *has* does not move from its S-internal position.

(31) a. $_{Comp}[who_i]_i$ [$e_i$ has read the book]
b. $_{Comp}[who_i\ did_j]_i$ [$e_i$ $e_j$ read the book]
c. $_{Comp}[what_i\ did_j]$ [Jay $e_j$ read $e_i$]

Since English does not have underlying object–verb order, there is no reason to require *has* to move. In fact, the nonoccurrence of **who did read the book* (where *did* is not stressed) suggests that an auxiliary verb may not move toward the front in such structures; the structure would have to be (31.b), because the unstressed *do* only occurs nonadjacent to the main verb. The nonoccurrence of (31.b) can be attributed to the fact that the empty subject fails to be governed by a coindexed Comp because of the presence of *did*. Although (31.c) does occur, here the trace of *what* is governed by the verb *read* and is coindexed with it (see note 11).[12] So, on hearing simple sentences like *wie heeft het boek gelezen?* and analyzing them as (30), the Dutch child relaxes the restriction on government, such that the empty subject is properly governed. This entails the well-formedness of complex sentences like (29). The English-speaking child hears no equivalent main clauses which motivate such a relaxation and therefore does not use structures like (28.b). Consequently, given a UG condition that empty items at PF must be properly governed, a degree-0 learnability account can be provided for the striking differences between Dutch and English embedded clauses.

This last case demonstrates that some properties of Comp which crucially affect the well-formedness of embedded clauses can be derived from main clause properties, given certain formulations of UG. However, this is not to say that all properties of Comp can be so derived. So, for example, English, unlike Dutch, French, German, Italian, Spanish, and many other languages, allows an element in Comp to be deleted, subject to certain conditions (primarily the UG condition already mentioned, namely, that empty elements at PF, including empty items and deletion sites in Comp, must be properly governed; see Weinberg et al. 1987): (32).

(32) a. Jay said (that) Kay left
b. Jan zei *(dat) Marie vertrokken was
c. Jean a dit *(que) Marie est partie

Similarly, French has a process whereby the complementizer *que* takes on the form *qui* if it governs an empty subject in an embedded clause (33).

(33) a. qui$_i$ crois-tu $_{S'}[_{Comp}$[qui/*que]$_S$[$e_i$ est parti] ]
who do you think that has left
b. qui$_i$ crois-tu $_{S'}[_{Comp}$[que/*qui]$_S$[ Marie a vu $e_i$] ]
who do you think that Marie saw

I cannot see how main clause data could motivate these processes, nor the marked device in English whereby the embedded subject of a nonfinite verb may be governed and assigned case by the higher verb: *I expect [her to win]*. A higher verb does not govern a lower subject unless the maximal projection (S′) dominating the lower clause is removed, on the assumption that *a* governs *b* only if they share all maximal projections (Chomsky 1981 and (27) above). The removal of this projection (often referred to as S′ deletion) is a marked process found in certain grammars and triggered by sentences like *I expect Jay to be happy, Jay was expected to be happy.*

Such phenomena require that children have access to at least the front of an embedded clause in order to set some parameters. This would suggest that the notion of degree-0 needs to be amended to 0-plus-a-little, as noted earlier. However, this messiness can be avoided if we think not in terms of clauses but rather of binding Domains (see 26). A binding Domain is defined in terms of a SUBJECT. The Domain for an element $x$ is the first clause or NP which contains an accessible SUBJECT. A SUBJECT is a higher NP or AGR, and a SUBJECT is accessible to $x$ if assigning the SUBJECT's index to $x$ violates neither the "i-within-i" condition (not relevant here) nor condition (26.c). (This will be illustrated in section 5.3.) The fixing of some grammatical parameters depends on access to a lower Comp and the subject of an infinitival, as noted. For each of these elements, the Domain in which there is a higher SUBJECT can only be the next higher clause and not the local clause, assuming as usual that an infinitival clause lacks the agreement marker. So in a structure like (34), if the lower clause lacks AGR, the only higher SUBJECT for the embedded subject is an element in the matrix clause. Similarly for Comp, regardless of whether there is AGR in the lower clause, because a lower AGR would not be higher than the Comp.

(34) $_{S'}$[ . . . NP AGR . . . $_{S'}$[Comp $_S$[NP . . .]]]]

If this is correct, it seems plausible to argue that some current formulations of UG allow us to maintain a strict version of degree-0 learnability. The child's triggering experience seems to consist only of robust positive data which may be analyzed within the framework of UG, given already established parameters of the particular grammar. Furthermore it is restricted to data occurring in an unembedded binding Domain, and we may claim that there is nothing new to be learned from embedded Domains. This permits a degree-0 account of a parametric difference between English and Chinese.

**5.3. Chinese AGR.** Aoun (1986), reanalyzing work by Huang (1982), argues that Chinese (whose verbs do not vary morphologically depending on the choice of subject) lacks AGR in INFL on the basis of sentences like (35):

(35) zhangsan$_i$ shuo $_{S'}$[ziji$_i$ hui lai]
Zhangsan say self can come
"Zhangsan said that himself will come"

Such sentences indicate that the binding Domain for the anaphor *ziji* is the matrix clause; this in turn entails that there is no accessible SUBJECT, that is, no AGR in the embedded clause containing *ziji*. In the English analogue to (35) (see the gloss) AGR would occur in the embedded clause and would act as an accessible SUBJECT for *himself;* consequently, the embedded clause would be the binding Domain for *himself* but would fail to be bound in that Domain, hence violating (26.a).

Datum (35) does not seem to be particularly exotic. It is also consistent with degree-0 learnability, if we define degree-0 in terms of binding Domain at LF: The Domain for *ziji* is not an embedded one and therefore *ziji* is available to a language learner having access only to unembedded material.

Defining the parametric difference between Chinese and English so that Chinese lacks AGR accounts for several superficial differences, notably differences in sub-

ject–object asymmetries and in the interpretability of the adjunct "why." Let us see how this works. All subject NPs are coindexed with their AGR, under the usual subject–verb agreement. Consequently, in (28.a) the lowest *e* has no accessible SUBJECT. The local candidates are *Kay* and AGR, but in each case assigning the index *k* to the *e* would entail that *e* would be coindexed with an argument position, that of *Kay;* this would violate (26.c). The same holds for the higher AGR and *Jay*. Therefore, in (28.a), *e* has no accessible SUBJECT, hence no Domain, and so the binding theory does not hold for this element and it has no locality restriction. Consequently, direct objects may move long distances generally. But now consider an empty element in subject position (28.b). Here *e* does have an accessible SUBJECT; assigning it the index of the local AGR does not entail a coindexing with some other argument position; AGR is therefore "accessible." As a result, *e* must be coindexed with a higher element in its local clause in order to satisfy (26.a). Since it has no local antecedent, the structure violates the binding theory (in addition to the PF condition of proper government, see section 5.2). Now compare this with (28.c), which lacks an overt complementizer at the front of the embedded clause. Again the empty subject in the embedded clause has an accessible SUBJECT for the same reason as in (28.b), but here there is a local antecedent, namely the indexed Comp, and condition (26.a) is satisfied.[13] This analysis provides a way of accounting for this kind of subject–object asymmetry, whereby extraction of a wh-word from object position (whether taking place in the syntax or in logical form) is quite free, but extraction from subject position is sensitive to the presence of a complementizer.

Although such asymmetries are widespread in the languages of the world, they do not occur in Chinese, as pointed out by Huang (1982). Saying that Chinese lacks AGR provides an explanation. In general, Chinese has its interrogative words occurring not at the front of the sentence but in the position in which they must be understood, that is, "in situ." Most languages allow this possibility (cf. English *I wonder who bought what*, where *what* remains in situ) but Chinese has no general alternative of syntactic movement. For such constructions the relevant scope relations are usually expressed through a process of "movement" in logical form (LF), the level of representation where quantifier-variable binding is expressed. A sentence like (36) is ambiguous and may have the LF representations of (37.a) or (37.b).

(36) ni xiang-zhidao$_{S'}$[shei mai-le sheme]
you wonder who buy-ASP what
a. "what is the x such that you wonder who bought x?"
b. "who is the x such that you wonder what x bought?"

(37) a. $_{S'}$[sheme$_j$ $_S$[ni xiang-zhidao $_{S'}$[shei$_i$ $_S$[e$_i$ mai-le e$_j$] ] ] ]
what you wonder who buy-ASP
b. $_{S'}$[shei$_i$ $_S$[ni xiang-zhidao $_{S'}$[sheme$_j$ $_S$[e$_i$ mai-le e$_j$] ] ] ]
who you wonder what buy-ASP

Because Chinese lacks AGR, neither *shei* nor *sheme* has a binding Domain and therefore both may move long distance in LF. In (37.a,b) $e_i$ has no accessible SUBJECT; there is no AGR and *ni* is not accessible. Also, the object $e_j$ has no accessible SUBJECT for the same reason as the object in (28.a). So, lack of AGR in Chinese entails no subject–object asymmetries analogous to those of (28).

However, (38) is not ambiguous in the same way as (36). "What" may have wide scope over the matrix clause, with "why" having narrow scope over the embedded clause under "wonder," but not vice versa.

(38) ni xiang-zhidao$_{S'}$ [Lisi weisheme mai-le sheme]
you wonder Lisi why buy-ASP what
"what is the x such that you wonder why Lisi bought x?"

Here *sheme* has no accessible SUBJECT and thus no binding Domain, for the same reason that *who* had no Domain in (28.a): The candidate SUBJECTs are *Lisi* and *ni*, but coindexing with these elements would lead to a violation of condition (26.c), that a referential expression be argument free. Hence the binding theory is irrelevant for *sheme* and there is no locality restriction. *Weisheme*, however, is not a referential expression and therefore is not subject to condition (26.c); *Lisi* is accordingly accessible to the trace of "why" and the embedded clause is its binding Domain. Hence (39.a) is a possible LF for (38), but not (39.b).

(39) a. $_{S'}$[sheme$_j$ $_S$[ni xiang-zhidao $_{S'}$[weisheme$_i$ $_S$[Lisi e$_i$ mai-le e$_j$]]]]
b. $_{S'}$[weisheme$_i$ $_S$[ni xiang-zhidao $_{S'}$ [sheme$_j$ $_S$[Lisi e$_i$ mai-le e$_j$]]]]

It is a general property of "why" and "how" that they are not referential expressions, and thus not subject to (26.c). Hence the contrast in English between (40.a,b); similarly in main clauses (41.a,b).

(40) a. Jay wondered who lives where
b. *Jay wondered who left why
(41) a. who lives where?
b. *who is leaving why?
(42) a. $_{S'}$[who$_i$]$_i$ $_S$[e$_i$ AGR$_i$ is leaving why]
b. $_{S'}$[who$_i$ why$_j$]$_i$ $_S$[e$_i$ AGR$_i$ is leaving e$_j$]

The structure of (41.b) is (42.a) and the LF is (42.b). When *why* moves to Comp at LF, its trace, $e_j$ has an accessible SUBJECT and thus a binding Domain, but fails to have an antecedent high enough in that Domain. The Comp must bear the index of the subject $e_i$ because otherwise the empty subject, for which AGR is an accessible SUBJECT, would fail to be bound in its Domain. The identical structure with *where* instead of *why* is well-formed because *where* is a referential expression and therefore has no Domain and is not subject to the demands of the binding theory (26). Consequently, saying that Chinese lacks AGR accounts for many puzzling differences between English and Chinese.

Now, although (41.b) does not occur in English the corresponding sentence is well-formed in Chinese.

(43) a. shei weisheme mai-le shu
who why buy-ASP books
"who buys books why?"
b. $_{S'}$[ [shei$_i$ weisheme$_j$]$_j$ $_S$[e$_i$ e$_j$ mai-le shu] ]

The LF of (43.a) is (43.b). Note that $e_j$ has an accessible SUBJECT, namely $e_i$, and must therefore be bound in its Domain, as it is. On the other hand, $e_i$ has no Domain, there being no AGR in Chinese; it therefore needs no local binder. Hence the distinction between (41.b), which is ungrammatical in English, and (43.a), which has the well-formed LF (43.b) in Chinese, lies in the fact that Chinese lacks AGR. Consequently, (43.a), involving only

material in an unembedded binding Domain, may also (like 35) act as a trigger for the correct parametric value: that Chinese lacks AGR.

This case has involved some theory-internal claims, as one should expect by now. My point throughout this target article has been that one's assumptions about the nature of the trigger will affect the details of one's technical analyses, including claims about both UG and particular grammars. In this example, the UG locality conditions formulated by Weinberg et al. (1987) enable us to establish a single parameter–the presence/absence of AGR–which accounts for many superficial differences between English and Chinese and which can be set on the basis of unembedded data, contrary to initial appearances. But, again, there is no logical relationship between the existence of locality restrictions and degree-0 learnability; the Chinese processes just discussed were quite local but it remained an empirical question whether there were parameters to be fixed which required access to complex sentences.

**5.4. Italian again.** If one thinks in terms of unembedded binding Domains rather than unembedded clauses as the basis for language acquisition, then a solution to the Italian problem (above) becomes available, as Guglielmo Cinque has pointed out (personal communication). Recall that Rizzi and Sportiche argued that S′ but not S is a bounding node in Italian and French. We showed how the French child could derive this conclusion from simple data like (24). However, it is not clear that comparable data are available to the Italian child. Analyses vary and (44.a) might be treated as (44.b,c, or d); it is not obvious that *quante* is extracted from a NP.

(44) a. quante ne hai viste?
"how many have (you) seen of them?"
b. $_{NP}[\text{quante } e_i]_k$ ne$_i$ hai viste $_{NP}[e_k]$
c. $_{QP}[\text{quante}]_k$ ne$_i$ hai viste $_{NP}[e_k\ e_i]$
d. $_{QP}[\text{quante}]_k$ ne$_i$ hai viste $[e_k]\ [e_i]$

However, if (45.a,b) are well-formed structures for the corresponding sentences, then movement has taken place across a NP and S, showing that S cannot be a bounding node. For arguments that such "small clauses" are dominated by S and not by S′, see Hornstein and Lightfoot (1987). In such a structure, the NP indicated, being a subject, cannot be reanalyzed. Furthermore, its binding Domain is the matrix clause since the lower S contains no SUBJECT (AGR).

(45) a. ne$_i$ ho visti $_S[_{NP}[\text{molti } e_i]$ correrglì incontro]
"of them (I) saw many run toward him"
b. ne$_i$ ho visto $_S[_{NP}[\text{il volto } e_i]$ sbiancarsi]
"of him (I) saw the face become white"
c. ne$_i$ ho visti $_S[_{NP}[\text{molti } e_i]$ affaticati]
"of them (I) saw many tired"

Hence simple, degree-0 data are available to make S into a nonbounding node in the grammars of Italian speakers. Although simple, the data are nonetheless slightly exotic and probably not of frequent occurrence in any child's experience. It is consequently not surprising that there is variability among Italian speakers in the setting of this parameter.

## 6. Conclusion

The viability of this hypothesis, that primary linguistic data are restricted to data occurring in an unembedded binding Domain, will depend on further case studies of the type considered here. I have examined some crucial cases which would seem to have motivated richer and more extensive triggers, and I have shown that, given certain formulations of UG, the relevant parameters can be fixed on the basis of data in nonembedded Domains. Further cases may be less tractable, of course, and this deliberately strong hypothesis may have to be revised and relaxed. The strong hypothesis directs attention toward discrepancies between phenomena occurring in embedded and unembedded Domains, but such discrepancies raise no particular learnability questions if embedded Domains may be part of the trigger experience.[14]

If something along these lines proves to be correct, it will explain the old observation that "what happens downstairs also happens upstairs, but not vice versa" (cf. Ross 1973). It will also have many consequences for the study of language acquisition under various conditions. To take one example, it will explain the rapidity of the creolization process. If most aspects of a person's grammar are triggered by simple data, the creole-speaking child who has an impoverished input is under no great disadvantage in developing a normal mature capacity. As a result, there is no need to postulate that creole grammars manifest the linguistic genotype in some more or less privileged way (Bickerton 1984), and thus one can avoid the unacceptable consequences of that claim (Lightfoot 1988).

However, there is a more fundamental point: An intrinsic relationship exists between claims made about the three entities of (1); and any claim made about, say, UG entails certain assumptions about the trigger. Those assumptions should be spelled out because they may vary in their plausibility, with consequences for the claim about UG and the particular grammars under investigation. Given the richness of current theories and the productivity of comparative work on parametric variation, claims about the trigger can be used as one basis for evaluating hypotheses. In fact, ignoring the trigger at this stage runs the risk that one's hypothesis may be off the mark. I have tackled the matter of the trigger experience here very much from a linguist's perspective, being concerned with consequences for claims about UG. However, the matter relates to many aspects of the acquisition process not touched in this article, and I hope that other perspectives may be able to tell us more about the true nature of the trigger.[15]

ACKNOWLEDGMENTS

Thanks to Joseph Aoun, Guglielmo Cinque, Bob Freidin, Norbert Hornstein, Jim McCawley, and Luigi Rizzi for helpful comments on an earlier draft of this paper, and to audiences at Princeton and the universities of Florida, Minnesota, North Carolina, Toronto, and Southern California.

NOTES

1. (2.b) provides only one possible structure for a noun phrase consisting of a head noun followed by a prepositional phrase, whereas (2.a) provides more than one structure: *Student from New York* can (and must) have the structure $_{N'}[_{N'}[_N[\text{stu-}$

dent]] $_{PP}$[from New York]], whereas *student of physics* is $_{N'}$[$_{N}$[student] $_{PP}$[of physics]]. The pronoun *one* refers to a preceding N'. *Student* is a N' in (4.b), hence a referent for *one*, but not in (4.a). In (5.a) both *suit* and *old suit* are instances of N' and thus possible referents for *one*, hence the ambiguity. For details and the reasons why *student from New York* and *student of physics* must have different structures, see Lightfoot (1982).

**2.** The rules of (2.a) allow an adjective to occur before a N or N', but convention (6) says nothing about possible adjective positions. This is appropriate because convention (6) stipulates the basic geometry of categorial structures; this basic geometry may be supplemented in response to experience. The position of adjectives is fully determined by a normal child's experience; exposure to phrases like *the tall student* will suffice to show that an English N' may have an adjective at its front. There is no poverty-of-stimulus problem for such adjectives which would require a richer convention in UG, as far as I can see.

**3.** Young children are known to have great difficulty in detecting for themselves the absence of forms, even when confronted with carefully prepared paradigmatic sets of patterns (Sainsbury 1971; 1973).

**4.** The notion of "markedness" has led to much confused discussion. UG includes a theory of markedness which leads one parameter setting to be preferred over another and permits "core grammar" to be extended to a marked periphery (Chomsky 1981, p.8). So the unmarked parameter setting is adopted in the absence of contrary evidence, but specific evidence will be required for a marked setting.

**5.** Baker (1979) discusses a transformational movement rule relating *John gave the book to Alice* and *John gave Alice the book*, which does not generalize to *report* and *say*. The fact that the rule is not entirely general suggests that negative data are needed to establish the limits to the generalization. He went on to show that a lexical relationship is preferable to a movement analysis and circumvents the apparent learnability problem if children are conservative in establishing the lexical properties of verbs, generalizing only within narrowly prescribed limits. However, children are not entirely conservative in this regard and they do, in fact, overgeneralize double objects to certain verbs. For further discussion, see Mazurkewich and White (1984) and Randall (1985).

**6.** For example, I assume that the morphological case system had already become dysfunctional. It is quite possible that a different reanalysis would have taken place if the case system had been rich enough and robust enough to force children to interpret the preverbal NP as a direct object.

This change in the use of *like*, etc., has given rise to much recent discussion: see Allen (1986), Anderson (1986), Elmer (1981), Fischer and van der Leek (1983), Warner (1983). Alternative analyses are possible (see Lightfoot, forthcoming).

**7.** Sentences like *Kirsten has* $_{NP}$*[a blue cup] and Heidi has a red one* show that *one* must also refer to something smaller than a NP.

**8.** Wexler and Culicover (1980) use a particular kind of error detection procedure which effectively adds an extra level of sentence embedding to whatever triggering experience is needed. They argue that the relevant information is contained in structures with one level of embedding, but their error detection procedure then requires another level of embedding to reveal possible errors. This is discussed interestingly by Morgan (1986), who adopts a different error detection procedure and offers a degree-1 learnability proof (based on the assumption that children receive as input not only Wexler and Culicover's $[b,s]$ pairs but also a surface string with constituent structure assigned).

However, the error detection procedure reflects a more fundamental difference in orientation. Wexler and Culicover were not much concerned with variation in grammars and used a selection model of acquisition wherein the child tests various hypotheses in accordance with a specific evaluation metric. I have adopted the recent parameter-setting model, however, in which UG contains rich information about, say, possible phrase structure rules and specifies certain option-points which are set on exposure to particular data-sets. The issues I raise in the following three case studies would also be important for a selection model of acquisition not making rich assumptions about parameters in UG. A comparison of the two models, however, would be extremely complex and will not be undertaken here, since selection models have not addressed the kind of language variation discussed.

**9.** If there are grammars which generally prevent extraction from an embedded clause (as claimed for Russian), then there is a setting for this parameter which is less marked than the English value: Both S and S' are bounding nodes. This would block the second movement in (i), from the embedded Comp into the higher Comp.

(i) who$_i$ $_{S}$[did Jay say $_{S'}$[e$_i$ [Kay saw e$_i$]]

The English-speaking child is confronted with specific evidence that the default value does not hold, i.e., sentences corresponding to (i). The Italian child is driven to a still more marked value when confronted by (21.a).

Since direct questions like *quel livre Jean sait à qui offrir?* ("which book does John know to whom to offer?") occur in French (but not in Italian), a trigger with one level of embedding can be specified.

For more discussion of the Italian case and alternative analyses, see Adams (1984) and Grimshaw (1986).

**10.** Since (24) requires that S not be a bounding node in French, structures like (23.b) will also be well-formed. Consequently, there is no positive motivation in the PLD for an English-style reanalysis rule yielding (23.c). This is the right result because French shows none of the English-type phenomena that Hornstein and Weinberg used to motivate the reanalysis rule.

This kind of extraction is not possible from a subject NP (**de qui$_i$ est [[le frère e$_i$] peintre]*?) for reasons other than the Subjacency condition.

**11.** The Comp indexing convention permits Comp to take on the index of an element that it contains. An indexed Comp (but not an element inside Comp) is "high" enough to act as a proper governor at PF and as an antecedent for an anaphor at LF. See Weinberg et al. (1987) for details.

Rouveret and Vergnaud (1980) propose a coindexing relation between a NP and the item from which it receives a theta-role, and so a verb is coindexed with and thus properly governs its direct object. In exactly the same sense a Comp properly governs a coindexed subject in (28.c), (29), (30), etc.

**12.** In (31) I have shown the auxiliary verb moving to Comp. This may not be correct; it may move to some other pre-S position, perhaps adjoined to S, as in Lasnik (1981). Nothing hinges on this for present purposes. Similarly for the Dutch case (30).

**13.** This analysis assumes that a *wh* word moves first to its local Comp (28.c). This is required by the Subjacency condition if S is a bounding node in English (see above in the main text). So movement from the lower subject position to the higher Comp in (28.b) would cross two bounding nodes, in addition to violating the binding theory (26.a) and the PF condition that empty items be properly governed (see below).

**14.** For example, standard analyses claim that Dutch and German are basically subject-object-verb and that the verb moves into the position of a complementizer in main clauses but not in embedded clauses. A child can presumably learn from primary data that verbs occur in complementizer positions and are therefore *moved* there from another base-generated position. Then two questions arise: (a) How do simple, unembedded data show from where the verb has moved, and (b) how do

children know that verbs may not occur in complementizer positions in embedded clauses? For discussion bearing on the first of these questions, see Travis (1987). Lightfoot (forthcoming) discusses the acquisition of object-verb order in such "verb-second" languages.

15. This target article represents a chapter of a larger work in progress; there I consider historical change in language and the way in which new parameter settings are adopted by children at certain stages in the history of a language (Lightfoot, forthcoming).

# Open Peer Commentary

*Commentaries submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.*

## Some observations on degree of learnability

C. L. Baker

*Department of Linguistics, University of Texas at Austin, Austin, TX 78712*
**Electronic mail:** *liar454@iv1.cc.utexas.edu*

Lightfoot argues with great clarity and persuasiveness that the choice among grammatical theories depends not only on assumptions about the systems that are attained, but also on assumptions about the nature of the linguistic data available to language learners. A particularly important point is made in section 2, where he notes "a curious silence about the triggering experience" in discussions of linguistic theory in which one would expect to find potential triggering experiences discussed. Lightfoot goes on to say that anyone who seriously proposes a language-particular rule, lexical entry, or parameter setting should be able in principle to suggest a plausible kind of triggering experience which would suffice for its selection.

Going beyond these general methodological remarks, Lightfoot advances an extremely restrictive view of the "degree" of the data that is relevant for a child's acquisition of a native language. Specifically, he proposes that the boundary between degree-0 and degree-1 data represents an important dividing line between plausible and implausible linguistic triggering experience. My primary purpose in this commentary will be to argue against this thesis. I want to suggest instead that it would be wrong to exclude all positive degree-1 data, and also that it would be wrong *not* to exclude certain positive degree-0 data. As a final incidental point, I will use Lightfoot's logic to call into question the analysis of Chinese he proposes near the end of his article.

An obvious preliminary point to be made concerning degree-1 data is that there is no scarcity of it in the child's linguistic environment. In this respect, then, we do not have the same kind of argument for excluding degree-1 data as input to a theory that we have for excluding negative data. Thus, the actual situation in which a first language is learned does not in and of itself provide any compelling reason to avoid using data of degree greater than zero.

Granting that data of degree greater than zero are available in ample quantities, we might still ask whether they make a truly essential contribution to language learning. On this question, Lightfoot offers a number of detailed case studies in which he suggests potential degree-0 triggering data in situations that might have appeared to require data of higher degree. However, at the end of section 5.2, he goes on to discuss some cases that seem to warrant the relaxation of the degree-0 requirement to "degree-0-plus-a-little."

Several constructions not discussed in the target article suggest that even this more permissive thesis is too restrictive. For example, a language learner apparently needs to find out whether relative clauses in the language that he is learning call for gaps or whether they call for resumptive pronouns. Learners face similar problems with constructions like the following English result clause:

(1) John$_i$ is too tired [to invite _____$_i$ to your house for dinner].

In particular, the learner needs to get far enough into the embedded infinitival to determine that it can contain a nonsubject gap. Some languages do not permit such a construction for the case in which the matrix subject is coreferential with a nonsubject in the infinitival, allowing instead only a structure like that in (2):

(2) John$_i$ is too tired [to invite him$_i$ to your house for dinner].

Turning to the other side of the degre-1/degree-0 boundary, I would like to argue that not all degree-0 examples have equal plausibility as potential triggering experiences. Near the conclusion of his discussion of Aoun's account of Chinese, Lightfoot gives the following as a simple sentence that would suffice to show (under Aoun's assumptions) that Chinese lacks AGR:

(3) (Lightfoot's (43)) shei weisheme mai-le shu
who why buy-ASP books
"who buys books why?"

Even though this is a degree-0 sentence, its potential usefulness as a triggering datum for a "no-AGR" parameter setting seems quite doubtful to me. I base this skepticism on the hunch that questions of this type are not significantly more common in causal Chinese conversation than they are in casual English conversation. There is a more general point to be made here. Whereas it is now widely recognized that most negative data about a language are best viewed as "nonprimary" in nature, it is less often appreciated that at least some positive data need to be accorded the same status. That is, certain types of acceptable sentences occur so rarely if at all in normal language use that their acceptability is more properly viewed as a result of language acquisition than as an input to it. (Many of the standard *pictures of each other* examples from English almost certainly have the same status.)

Regarding the proposed analysis of Chinese, recent work by Susan Wilcoxon (1988) casts serious doubt on the validity of the conclusions that Huang and Aoun have drawn from the existence of sentences like Lightfoot's (35), in which *ziji* appears at first glance to be the subject of the subordinate clause:

(4) (Lightfoot's (35)) zhangsan$_i$ shuo $_{S'}$[ziji$_i$ hui lai]
Zhangsan say self can come
"Zhangsan said that himself will come"

Wilcoxon shows that as well as serving as a reflexive pronoun *ziji* has an added use as an optional intensive adverb that can be linked to subjects of all kinds. When *ziji* is used with a subject that happens to be empty, the following result is predicted:

(5) zhangsan$_i$ shuo $_{S'}$[$_{NP}$[e$_i$] $_{Adv}$[ziji$_i$] hui lai]
Zhangsan say he$_i$ self can come
"Zhangsan said that he himself will come"

The reactions of Wilcoxon's informants suggest that the intensive interpretation is the only one the sentence in question allows. Besides removing one piece of evidence for postulating this parametric difference between English and Chinese, Wilcoxon's reanalysis makes such sentences ineffective as "de-

gree-0-plus-a-little" triggering data for the parametric setting in question.

If we now take seriously the methodological remarks in section 2, we are forced to question the validity of the proposed parametric analysis of Chinese. Two possible triggering data were offered for the postulated parametric setting. The "degree-0-plus-a-little" datum was shown to be deficient by virtue of its failure to force this parameter setting unambiguously, whereas the degree-0 datum was argued to be deficient by virtue of the extremely small chance that Chinese language learners would have of actually hearing it uttered. Unless triggering data can be found that do not suffer from either of these deficiencies, we have grounds for doubting that English and Chinese really differ in the way proposed.

## On triggers

Hugh W. Buckingham

*Program in Linguistics, Louisiana State University, Baton Rouge, LA 70803*

In this commentary I would like to explore some thoughts that came to mind in reading Lightfoot's target article. The first has to do with the very impressionistic notion of "exotic" sentences or situations in which sentences arise. Granted that if children must await degree-2 sentence types or if they must hold off setting a parameter for n-bar until they find themselves in a situation concerning "old suits" or the like, then their syntactic development will be delayed, but we still do not have anything like a technical definition of the "exotic" sentence or situation. Eventually, some sort of metric for this will have to be developed, but at present I am not sure what it would be.

Second, it has often been remarked that children have little if any notion of grammaticality. Consequently, it is still unclear to me just how a child will reap the benefits from this or that crucial trigger structure. It would seem to me that if a trigger is to have any sort of effect whatsoever, then the child will somehow have to know that the trigger is grammatical. How could the fact that a trigger is acted upon to set a parameter ever be empirically demonstrated? How would, or could, a test be formulated to demonstrate the utility of a trigger?

My third point hinges on the fact that parameter settings have a sort of typological ripple effect. Setting the parameter, for instance, for [+pro-drop] (pronoun dropping), will have a cascading effect–switching levers that will allow the possibility of postverbal subjects, the possibility of object–verb structures, and the like (e.g. Phinney 1987; Torrego 1984; White 1987). The question then becomes that of establishing the necessary type of trigger. Would the trigger need to be some sentence without an overt subject, or would a verb–subject (VS) or object–verb (OV) structure work equally well? Even if the initial setting (the unmarked case) would be (due to the "subset" principle) at [−pro-drop], how would we establish which trigger type would be more crucial for a pro-drop language? A trigger structure with null subject? A VS structure? An OV structure? Or any one of them? The answer is not clear to me, but the typological nature of the theory of parameter settings makes this question an important one to consider.

The fourth issue I would raise is that before any notion of proper and/or necessary triggers can be put forth clearly, we will have to have a hold on just what the proper parameters are for a specific language. Most studies with which I am familiar specify that, for English, NP and S are the bounding nodes; but recently Grimshaw (1986) has made a case for considering S-bar the bounding node, not S. She makes an important distinction between initial movement of alpha and subsequent movements of alpha. For her arguments to go through, we must agree upon the "essential" (Grimshaw 1986, p. 366) grammaticality of sentences such as:

(a) [[ Which book$_i$ ] [$_S$ did the students forget [$_s$[who$_j$] [$_s$ t$_j$ wrote t$_i$]]]]

(b) [[ Which book$_i$ ][$_s$ did the T.A.'s tell the students [$_s$[that] [$_s$they shouldn't forget [$_s$[who$_j$] [$_s$ t$_j$ wrote t$_i$]]]]]]

One Comp can be skipped on the initial movement on Grimshaw's view, but note that two S nodes are crossed in both (a) and (b). If S is a bounding node in English, this situation should produce ungrammatical sentences. Again, the argument only goes through if we agree on the grammatical status of (a) and (b). As Grimshaw points out, however, (a) and (b) are certainly better than sentences such as:

(c) [[ Which book$_i$ ][$_s$ did the students forget [$_s$[who$_j$] [$_s$ t$_j$ told them [$_s$[that] [$_s$ Noam Chomsky wrote t$_i$]]]]]]

The initial movement poses no problem in (c), but as Grimshaw suggests, subsequent movements may not skip a Comp, which is exactly what must happen in (c). In this case, the moved item crosses two Ss and two S-bars. Since the "essential" grammaticality of sentences (a) and (b) presumably demonstrates that S is not a bounding node, the ungrammaticality of (c) would force the analysis that S-bar is indeed the bounding node for English. Grammaticality judgments aside, several problems arise for the theory of triggers here. First, the sentences that crucially bear on these issues are at least of degree-2 complexity. Linguistic theory, however, must be carried on at this level. The theory of triggers would be in deep trouble if it turned out that degree-0 structures would lead the English-speaking child to set the bounding parameter at S, but that later complex sentences of degree-2 or more would force a resetting to the S-bar. In general, though, my point is that issues such as the ones Grimshaw raises must be settled before any sense can be made of the nature of bounding node triggers–degree-2 or degree-0.

Although I previously mentioned the difficulty of demonstrating that the child would have an easy time of it appreciating the actual utility of positive trigger structures, Lightfoot and all language learning theorists are certainly right in emphasizing the uselessness of negative data, that is, "String not in language, Johnny." Even in the days of the evaluation metric view of acquisition, negative data (ungrammatical sentences) were formally proven to be inconsequential for developing grammatical knowledge. The formal existence of weakly equivalent grammars demonstrated clearly that deciding between grammars could not involve negative data. The problem there, of course, was that positive data could be of no help either, since the two grammars generated the *same* set of strings. Theories of simple phrase structure grammars told us as much. For this reason, the onus was placed on some sort of "metric" to help decide between two grammars. Considering strings (grammatical or ungrammatical) alone was no help at all. As Lightfoot points out, however, the parameter-setting model has replaced the selectional model that required the aid of an evaluation metric. The latter model, which grew out of early Chomskyan thinking, simply did not produce results. Williams (1987, p. vii) puts it this way, "The parametrized model of grammar grew out of the failure, on the part of grammarians, to make any progress in understanding grammar selection with this 'evaluation metric.'"

Negative evidence is out on both views, but now, with the model of parametrized grammar, positive evidence is utilizable in the form of triggers. The question then becomes one of deciding just how a child will appreciate the positive evidence *as* positive evidence and of deciding the degree of complexity of the positive evidence as well as the exact structures of the positive evidence. Lightfoot appears to be on the right track here in seeking the maximum information in degree-0 struc-

tures. Finally, linguistic theory will have the responsibility of driving the whole enterprise, since without knowing the principles and parameters for Universal Grammar and how the parameters are set for specific languages in question, we cannot know which triggers will be required to learn those languages.

## Parameter setting in "instantaneous" and real-time acquisition

Guglielmo Cinque

*Seminario di Linguistica, University of Venice, 30124 Venice, Italy*

Much as in equations with two constants and one "unknown," whenever we have a reasonable understanding of the grammatical knowledge attained and of the triggering experience, we can infer something about the contribution of UG. "Poverty-of-stimulus" arguments, such as the one about *one*-pronominalization discussed by Lightfoot, typically have this character.

But there is a class of cases in which we can gain independent access to properties of UG; that is, when we compare different grammars (the recognition that languages differ in either allowing or not allowing null subjects, for example, requires UG to have a specific two-valued parameter to that effect). In these particular cases, then, by treating UG and attained knowledge as known, we can treat the triggering experience as the "unknown," trying to determine with better accuracy what the general character of the trigger must be like for the child to arrive at the correct value of the parameter which yields the grammar.

The main goal of Lightfoot's target article is to determine in this way the character of the triggering experience beyond the little that is generally assumed: the unavailability of explicit instruction and direct negative evidence.

I think Lightfoot succeeds in defending the desirability and correctness of "degree-0" learnability in many controversial cases which would seem to require a more complex trigger. This is a result, which, in my opinion, one should be willing to abandon only if forced to do so.

I would like to raise a related point, one that is somewhat in the background in Lightfoot's discussion, but which may deserve some clarification. The problem of parameter setting is usually posed in the context of the simplifying assumption that acquisition is instantaneous (cf. Chomsky 1975, pp. 119–22). My point is that such an assumption may not be just a convenient (and innocuous) idealization. It may well prove to be the only pertinent way to approach the question. Considerations of the actual stages of acquisition (though interesting in their own right) run the risk of distorting matters. Let me consider the pro-drop (pronoun-dropping) parameter as a case in point.

Rizzi (1982) suggested, on the assumption that negative evidence is not available to the child, that the English value of the parameter should be the least marked value. This is because, starting from the wrong setting, the Italian value can be learned from extremely simple positive data contradicting the other setting, whereas the English value cannot.

Note that, even though the two values do not give rise to languages that stand in a proper subset relation to each other, as noted by Hyams (1986), the kind of data that could act as a trigger to reset the parameter from the Italian to the English value (e.g., the existence of lexical expletives) is radically limited.

There is a complication, however. Hyams (1986) noted that even earlier stages of non-pro-drop languages are pro-drop, so that the Italian value must apparently be taken as the "initial" setting of the parameter (whether or not it is also the least marked; cf. Rizzi 1986, fn.27). This disrupts the rather straightforward picture seen above, raising the problem of what the evidence is that forces the English child to reset the parameter from the larger to the smaller language.

Hyams (1986) suggests, as noted, that the existence of lexical expletives and pragmatically neutral lexical pronouns (not found in pro-drop languages) provides the required positive evidence to reset the parameter to the English value. But some questions remain: Why should the Italian value be the initial setting? Why should the resetting occur at the particular time of development that it occurs? Hyams's answers, based on her isomorphic principle, though interesting, are not entirely convincing. In particular, one might expect that different children reset the parameter at (widely) different times (depending on such variables as exposure to relevant data, intelligence, etc.). Similar questions arise within the solutions to this problem suggested by Weinberg (1987) and by Lebeaux (1987).

Recent work on early child grammar by Andrew Radford opens the way to a more radical alternative, one that eliminates the need to consider the Italian value as the "initial" setting and renders real-time acquisition indistinguishable from instantaneous acquisition for the problem at hand. Radford (1988) suggests that several apparently unrelated properties that distinguish the grammar of English children up to the mean age of 22 months from the grammars of later stages should be seen as different manifestations of a single, more abstract, property of the early grammar: the fact that the child's phrase structure has not yet developed the "nonlexical" projections of I(nflection) (IP = the traditional *S*) and (C)omplementizer (CP = the traditional *S'*).

Radford observes that if these two projections are missing (with their respective heads, I and C, and specifiers), then all phenomena which involve such positions should also be missing. In this way, he is able to relate several properties noted in the literature as typical of the child's grammar of this period, namely, that complementizers and "inverted" auxiliaries (both of which involve C) are missing; that no preposed wh-phrases (which move to the specifier of CP) are possible; that clauses lack modal auxiliaries and infinitival *to* (which occupy I); that verbs are not marked to *Tense* or *Agreement* (also found in I). (See Radford 1988 for a more careful discussion of these and further properties.)

According to this account, it is also to be expected that all such properties will emerge in strict succession, as indeed appears to be the case, and not at random intervals in time.

Note that among the missing positions is the structural subject of the clause (the specifier of IP). Now, this is the crucial fact for our pro-drop problem. Subjects, whether lexical or empty (when permitted by the pro-drop parameter), must be present in clauses, by the E(xtended) P(rojection) P(rinciple). Subjects of phrases other than clauses, on the other hand, need not be present (cf. *the destruction of the city*). But the EPP cannot be active until the structural subject position develops. So before that there is no requirement that a subject be present (when lexically present, it may either be a topic – cf. Gruber (1967) – or a "small clause" subject – cf. Radford 1988). All this implies that the setting of the pro-drop parameter does not even arise at this stage.

If so, in the pro-drop case (and perhaps in general) there is no *resetting* of the parameter, but rather *delayed setting*. Only when the required conditions mature does the setting occur; and at that point, it may apply as it does under the instantaneous acquisition model, with the English value as the least marked one, thus avoiding the problems of the alternative choice.

In this view, the instantaneous model represents abstractly how the parameter must be set at the relevant point of real-time acquisition. It is perhaps the only model giving the correct results, as seen, provided that we do not expect it to explain the actual course of acquisition, which may be affected by maturational factors (cf. Borer & Wexler 1987; Felix 1987).

## Causality and parameter setting

Robin Clark

*Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA 15213*

**Electronic mail:** *clark@beaune.ling.ucla.edu*

The discovery of the correct adult grammar seems quite tractable in a parameter-setting model. A parameter may be thought of as a linguistic property; for example, a property might be whether the object precedes or follows the verb in the target language. The learner would need only monitor the input data searching for examples that can be used to determine the correct parameter setting. If Universal Grammar specifies only a finite set of linguistic properties (parameters), then only a finite number of examples should be required for the learner to arrive at the correct target grammar. The goal of arriving at the correct knowledge state on the basis of finite input would then be achieved.

Lightfoot correctly observes that the nature of the input data must be carefully considered in developing a parametrized theory of Universal Grammar. Studies of "Motherese," for example, indicate that the primary linguistic data (PLD) consist largely of simple grammatical sentences.[1] If the data required to set a parameter are of such complexity that they are unlikely to occur in the PLD, then the probability that the learner will be able to converge on the adult grammar will decrease accordingly. A linguistic theory which purports to account for learnability using such a parameter would then be cast into doubt. A parameter which could only be set given a clause with three or four embedded clauses in it is unlikely ever to be set if only that the necessary input is so rare.

For the above reasons, Lightfoot's attempt to develop a "degree-0" model of parameter setting is on the right track. A "degree-0" theory would provide a plausible model of the linguistic environment that the learner must work in. I would argue, however, that the computational properties of the learner must be included in the theory along with Universal Grammar and the role of the PLD. In particular, a parameterized (selective) theory may still face some problems similar to those faced by inductive theories. Put briefly, given a piece of input data, how does the learner know which parameter to set? The learner must have some idea how the datum is relevant to the problem of arriving at the correct adult state.

To illustrate the type of problem faced by a parametrized learner, let us suppose that the learner is provided with a set of parameters by Universal Grammar. For each new input sentence, the learner attempts to assign that sentence a well-formed representation. If the learner succeeds, it does nothing to the current parameter setting and proceeds to the next input datum. If the learner fails, then it selects a parameter and resets it. The question is how the learner selects a parameter.

Consider, for example, "Exceptional Case Marking" (ECM) structures, as in (1):

(1) They believe [$_{\text{Clause}}$ each other to be ill]

The property of (1) that is of interest for present purposes is that the matrix verb, *believe,* stands in the structural relationship of government to the subject of the embedded infinitival clause, *each other*. Because of this structural relationship, the verb is able to assign abstract case to the subject of the infinitival. As a result, the subject of the infinitival is permitted to have phonological content (see Chomsky 1981a).

Not all languages have ECM structures, so their presence in a language must be the result of setting a parameter. Some languages allow for Case assignment to the subject of infinitivals without government by a verb like *believe;* modern Irish and Latin (Chung & McCloskey 1987) are examples of this kind of language. The latter type of language will have sentences very similar to the sentence in (1); the Irish example in (2) shows this:

(2) Is cuimhneach leo [$_{\text{Clause}}$ iad a bheith ar seachran]
[Copula] mindful with-them they to-be lost
"They remember being lost."

Despite the apparent similarity between (1) and (2) with respect to the phonologically overt subject of the infinitival clause, the two structures differ in ways that have interesting consequences for the resulting language. Because of the government relation, English allows anaphoric noun phrases (like *each other*) to occur in the subject position of the infinitival, whereas Irish does not because it lacks the appropriate government relation:

(3) *Shil siad [$_{\text{Clause}}$ a cheile a bheith breoite]
thought they each-other to-be ill
"They thought each other to be ill."

Clearly, setting the relevant parameters correctly is important to explaining the resultant grammaticality judgments.

Upon encountering examples like (1), how does a child in an English-speaking environment know that the language has ECM rather than a Case-marking system like Irish? One might imagine that the child would compare the predictions made by the two different parameter settings to decide which fits the facts better. Thus, the child could note that examples like (1) allow for an anaphoric noun phrase in the subject position of the infinitival clause and could then (correctly) select the setting that allows ECM. But the child could also incorrectly conclude that the language was like Irish in having special Case marking for infinitival subjects, and like Chinese (see the target article for discussion) in allowing for "long distance" anaphors. Such a language would permit examples like (1) without ECM. In fact, the language might well be a superset language of English; it would contain all the grammatical sentences as well an infinite number of ungrammatical strings. Assuming that the child encounters only simple, grammatical utterances (the underlying assumption of the "degree-0" framework), no data available to the child would overrule the bad hypothesis.[2] The learner must connect the distribution of anaphors with the ECM parameter and not assume that the two phenomena are unrelated.

The problem is that each parameter setting makes a wide variety of predictions about the form of the resulting language, and many of these predictions are made given that other parameters are set in a particular way. Parameter settings are not discrete but interact with other parameter settings to generate the adult language. The learner's problem in setting a parameter appears to be analogous to the "frame problem" familiar from artificial intelligence (McCarthy & Hayes 1969); how does one go about changing a knowledge state automatically in the face of a dynamic world? In the present context, how would the learner know when to stop generating and testing predictions made by possible parameter settings and combinations of parameter settings? It seems unlikely that an effective procedure can be given that will tell the learner that it can stop testing predictions and set a parameter. If no such procedure can be given, then the learner might never actually set a parameter.

Children are nevertheless quite successful at acquiring language quickly and efficiently. This suggests that the learner is so structured that the relevant parameter can be set given particular input data. Discovering how the learner has this built-in capacity will play a fundamental role in cracking the logical problem of language acquisition. Lightfoot is right in emphasizing the role of the input in language learnability; but the relationship between the input data and Universal Grammar is crucially mediated by a learner. The problem will not be solved until we can achieve an understanding of how the input data, Universal Grammar, and the form of the learner interact to arrive at the adult state of knowledge of language.

NOTES

1. On studies of Motherese, see Newport et al. (1977) and the references cited there. Wexler and Culicover (1980) provide an exten-

sive discussion and criticism of the assumptions underlying many studies of Motherese.

2. This situation violates the "Subset Principle." See Berwick (1985), Manzini and Wexler (1987), Wexler and Manzini (1987) for discussion.

## Zero-stimulation for parameter setting

Robert Freidin[a] and A. Carlos Quicoli[b]

*[a]Department of Philosophy/Program in Linguistics, Princeton University, Princeton, NJ 08544-5264 and [b]Department of Spanish and Portuguese (UCLA); Department of Romance Languages and Literatures/Program in Linguistics, Princeton University, Princeton, NJ 08544-5264*

**Electronic mail:** *[a]bob@confidence.princeton.edu and [b]cquicoli@confidence.princeton.edu*

The acquisition of a language involves two components: the contribution of the individual and the contribution of the linguistic environment. What cannot be attributed to the linguistic environment must be attributed to the individual. Within the framework of generative grammar, the contribution of the individual is represented as a system of rules[1] and a related system of principles which determine the form or application of the rules, or the representations the rules generate. These two systems constitute Universal Grammar (UG), an explicit theory of the biological endowment of our species (universal across the species) which allows a human being to acquire a language. It is further assumed that these systems (actually, subsystems of UG) may be realized differently for specific languages within certain limits. For language acquisition to occur, an individual must be exposed to (primary) linguistic data which allow him to acquire the grammar of a particular language. Lightfoot's target article focuses on the nature of the evidence required so that the child endowed with UG can develop the grammar of a specific language.

In line with recent advances in generative grammar, Lightfoot adopts the view that specific grammar develops from the fixing of certain open values (parameters) in the subsystems of UG. In this view the acquisition of language is largely a result of parameter setting on the basis of the linguistic evidence available to the child. Lightfoot's main new claim is that the evidence required for parameter setting may basically be limited to properties of main clauses (without reference to embedded clauses – what he calls degree-0 learnability). In this way, Lightfoot's target article constitutes a serious effort to clarify the relationship between the contribution of the individual's UG and of the linguistic environment. Although we find ourselves largely in agreement with Lightfoot's general approach, there are questions of detail that require further clarification and revision.

The term "parameter" in the literature of generative grammar (including Lightfoot's article) covers several distinct notions. Because this has a direct bearing on the issue of language acquisition in the parameter-setting model, it is necessary to distinguish between the different notions. Once the distinctions are made, it becomes clear that the child's triggering experience (i.e., the linguistic evidence the child uses in constructing a grammar) bears on one particular notion and not the others. We will argue that the evidence for UG based on the poverty of the stimulus is also evidence that certain parameters of UG are preset innately (i.e., in the absence of any triggering experience). This suggests that a certain sharpening of the parameter-setting model that Lightfoot's article explores is in order.

***1. On the notion "parameter of UG."*** Given a model of UG consisting of a subsystem of rules and a subsystem of principles, we can identify parts of the subsystems which are subject to variation across languages. The range of values involved in this variation constitute the "parameters" of UG. Thus we have two types of parameters: those associated with the subsystem of principles (*p-parameters*) and those associated with the subsystem of rules (*r-parameters*). For example, the values associated with the Subjacency principle (i.e., the choice of S or S′ as a bounding node, discussed in Lightfoot's section 5.1) would be instances of p-parameters. With respect to rules, the concrete values that are assigned to category variables in rule schema (e.g., Lightfoot's section 6) would constitute r-parameters. In Lightfoot's discussion (sect. 2, para. 6) the linear order of constituents also constitutes a parameter – and hence an r-parameter in our terms. One further set of r-parameters would be the possible category values assigned to α in the rule schema "move α" which represents the two distinct elementary transformational operations of substitution and adjunction.[2]

According to the parameter-setting model of language acquisition there is a crucial difference between p-parameters and r-parameters with respect to the role of linguistic experience. In the case of r-parameters, it is reasonable to assume that linguistic data are necessary and sufficient to fix values. Thus word order and the categories affected by movement rules will be determined primarily by data – the triggering experience. However, the setting of p-parameters cannot be explained in this way.

***2. P-parameters and the poverty of the stimulus.*** Principles of grammar such as Subjacency are motivated in part by empirical evidence concerning the ill-formedness of certain syntactic structures. For example, Subjacency in English (where the bounding nodes are taken to be S and NP) accounts for the following contrast.

(1) a. i. Who do you believe that Mary likes?
ii. $[_{S'}$ who$_i$ $[_S$ do you believe $[_{S'}$ $e_i$ that $[_S$ Mary likes $e_i$ ] ] ] ]
b. i. *Who do you believe the claim that Mary likes?
ii. $[_{S'}$ who$_i$ $[_S$ do you believe $[_{NP}$ the claim $[_{S'}$ $e_i$ that $[_S$ Mary likes $e_i$ ] ] ] ] ]

(In (1b.ii) and elsewhere the arrow indicates the connection which violates Subjacency.) Given standard assumptions, the wh-phrase in (1a) moves in a stepwise fashion through each intervening Comp position (the position between the S′ and S brackets in the structures above) and leaves a trace ($e_i$) in each position it moves from. In this analysis, (1a) does not violate Subjacency. In contrast, (1b) violates Subjacency because the movement of the wh-phrase to the front of the main clause crosses two bounding nodes (NP and S).

Since (1b) does not occur as part of the child's triggering experience, Subjacency must be assumed to develop in the absence of the relevant linguistic experience – what we shall call "zero-stimulation." Where it can be demonstrated that zero-stimulation holds with respect to a principle of grammar, that principle must be innate. This is a standard poverty-of-the-stimulus argument for establishing principles of UG.

If Subjacency has p-parameters (cf. Rizzi 1982, fn. 25; Chomsky 1981, p.52 ff.), then the argument from the poverty of the stimulus applies not only to the existence of the principle in UG, but also to the original setting of its parameters. The parameters for Subjacency must be preset in UG at the most restrictive settings. If not, then the language learner would require negative evidence (e.g. (1b)) to fix the p-parameters correctly.

This can be further illustrated by considering the case of wh-movement in Russian, which is more constrained than English (and Italian). In Russian, an interrogative pronoun (or wh-phrase) can be moved to the front of the its clause as in (2).

(2) a. Kavo ljubit Marija
who-ACC loves Mary-NOM
"who does Mary love?"
b. Ja znaju, kavo Marija ljubit
I know who-ACC Mary-NOM loves
"I know who(m) Mary loves."

(2b) illustrates that wh-movement applies inside an embedded clause just as it applies in a simple sentence like (2a). However,

a wh-phrase cannot be moved from an embedded clause to a matrix clause. Thus the Russian example corresponding to (1a) is ill-formed. Compare (1a) and (3).

(3) a. *Kavo govorit Ivan, čto Marija ljubit?
who-ACC says Ivan that Mary-NOM loves
"who does Ivan say that Mary loves?"
b. [$_{S'}$ Kavo$_i$ [$_{S'}$ govorit Kvan, [$_{S'}$ e$_i$ čto [$_S$ Marija ljubit e$_i$ ] ] ] ]

These facts can plausibly be explained by the Subjacency principle under the assumption that both S and S′ are bounding nodes in Russian. Since negative evidence such as (3) is not available to the language learner, we have another case of zero-stimulation. We are therefore led to assume that the p-parameters for Subjacency must be preset at S and S′ (and also NP, for reasons we won't discuss here).

If a child acquires a grammar of Russian, the p-parameters for Subjacency remain as preset in UG. But if the child is exposed to a language like English, positive evidence like (1a) will trigger a resetting of the p-parameters to allow the movement of a wh-phrase out of its clause. This is achieved by excluding either S or S′ as a bounding node for Subjacency. Again, negative evidence indicates that the correct choice for English is to exclude S′. The relevant example is given in (4).

(4) a. *Which books did John wonder whom Bill gave to?
b. [$_{S'}$ Which books$_i$ [$_S$ did John wonder [$_{S'}$ whom$_j$ [$_S$ Bill gave e$_i$ to e$_j$] ] ] ]

If S is a bounding node for English, then Subjacency correctly rules out (4a). If instead S′ is taken to be a bounding node, then (4a) is no longer excluded by Subjacency. The choice of S over S′ for English, which also depends on negative evidence, is yet another instance of zero-stimulation. By the logic of the argument concerning the initial setting of p-parameters, the resetting of p-parameters for Subjacency from {NP,S′,S} to {NP,S} must be innately specified as part of UG. Positive evidence like (1a) merely indicates that the p-parameters must be reset, but is not sufficient to determine the actual values.

A third situation arises in Italian and French, where examples analogous to (4a) are grammatical. (The following Italian examples are from Rizzi 1982, 18a-b, and the French examples are from Sportiche 1981, 45a-b.)

(5) a. Italian
i. Il mio primo libro, che crédo che tu sappia a chi ho dedicato, me e sempre stato molto caro.
"My first book, which I believe that you know to whom I dedicated, has always been very dear to me"
ii. Il mio primo libro,[$_{S'}$che$_i$ [$_S$credo[$_{S'}$e$_i$che[$_S$ tu sappia[$_{S'}$ a chi$_j$ [$_S$ ho dedicato e$_i$ e$_j$]]]]]] me e sempre stato molto caro.
b. French
i. Voilà quelqu'un à qui je crois que je sais lequel j'offrirais.
"Here is someone to whom I think that I know which one I will offer."
ii. Voilà quelqu'un[$_{S'}$ à qui$_j$[$_S$je crois[$_{S'}$e$_j$que[$_S$je sais [$_{S'}$lequel$_i$[$_S$ j'offrirais e$_i$ e$_j$]]]]]]

If S were a bounding node in Italian then the connection between the traces of the wh-phrase *che*$_i$ in (5a.ii) would be incorrectly excluded by Subjacency because it crosses two S nodes. If, however, S′ rather than S is a bounding node for Italian, as argued by Rizzi, then the connection between these traces crosses only one S′ and (5a.ii) is permitted (since no Subjacency violation occurs). An analogous argument has been made by Sportiche for the French example (5b). In contrast to (5), the examples in (6) show that Subjacency does hold for Italian and French when two S′ nodes are crossed.

(6) a. Italian
i. *Il mio primo libro, che so a chi credi che abbia dedicato, me e sempre stato molto caro.
"My first book, which I know to whom you believe that I dedicated, has always been very dear to me"
ii. Il mio primo libro,
[$_{S'}$ che$_i$ [$_S$ so [$_{S'}$ a chi$_j$ [$_S$ credi [$_{S'}$ e$_i$ che [$_S$ abbia dedicato e$_i$ e$_j$] ] ] ] ] ]
me e sempre stato molto caro.
b. French
i. *Voilà quelqu'un à qui je sais lequel je crois que j'offrirais.
"Here is someone to whom I know which one I think that I will offer."
ii. Voilà quelqu'un
[$_{S'}$ à qui$_j$ [$_S$ je sais [$_{S'}$ lequel$_i$ [$_S$ je crois [$_{S'}$ é$_j$ que [$_S$ j'offrirais e$_i$ e$_j$] ] ] ] ] ]

In (6a.ii) for example, while the connection between the wh-phrase *a chi*$_j$ and its trace crosses only one S′, the connection between *che*$_i$ and its closest trace crosses two S′ nodes in violation of Subjacency. This holds true for the French example. By the logic of the general argument, the resetting of the p-parameters for Subjacency in Italian and French (i.e. from {NP,S′,S} to {NP,S′} requires positive evidence – either examples like those in (5) above or simplex structures of the sort discussed by Lightfoot (see his examples 24 and 44).

What we have attempted to show here is that p-parameters in UG must be preset because there is no relevant triggering experience of any sort (including degree-0 data). By the logic of the argument from the poverty of the stimulus, the original setting is a case of zero-stimulation. The trigger experience is relevant only to the *resetting* of the p-parameters. And even for resetting, the change values in some cases (e.g. Subjacency for English) must be innately prespecified. If this is correct, then the trigger experience, which is important for the setting of r-parameters, does not determine the setting of the p-parameters of UG.

ACKNOWLEDGMENTS

We would like to thank Carlos Otero for helpful comments on an earlier draft of this commentary, and Charles Townsend for his help with the Russian data.

NOTES

1. We assume that the rule system of UG consists of abstract mechanisms (rule schema) which may be instantiated as language particular rules in specific grammars. On this view, such mechanisms would include the elementary transformational operations
i. adjoin α to β
ii. substitute α for β
iii. delete α
where α and β are syntactic category variables. This subsystem of UG does not include language particular phrase structure rules or transformations.

2. There is a third use of the term "parameter" in the study of crosslinguistic variation where it has yet to be established that any p- or r-parameter is involved. The most often cited case is the so-called null-subject parameter whereby some languages allow null subjects in finite clauses and others do not. The notions "null subject" versus "non-null subject" are not values of any subsystem of UG. This difference in phenomena across languages may arise as a result of parameter setting for other principles – e.g., the Empty Category Principle vis-à-vis the values for proper governors (see Chomsky 1981a) or for the direction of government (see Adams 1987).

## Infinitely nested Chinese "black boxes": Linguists and the search for Universal (innate) Grammar

Allen D. Grimshaw

*Department of Sociology, Indiana University, Bloomington, IN 47401*

***1. Introduction.*** As an outsider I have always admired and even envied linguists; at the same time I have sometimes found

myself sympathizing with them over the apparent intractability of some of the intellectual puzzles with which they must deal. Finally, I have been puzzled by at least some linguists' notions of what constitutes evidence for often intriguing theoretical claims. I admire linguists because they have taken on the challenging task of explaining language and because in doing this they have avoided a separation between "theory" and "research" which has inhibited the development of strong theory(ies) in my own discipline of sociology. I sympathize with linguists because they must confront difficult and heavily freighted and possibly unanswerable questions of semantics. I envy linguists because some of their data (phonological) are so easy to obtain and their measurement so crisp – and because working with those materials has permitted linguists to develop much more elegant "discovery procedures" than generally obtain in my own field. These positive reactions make it hard for me to understand why some of the most productive, thoughtful, and innovative linguists continue to rely primarily on their own "intuitions" (or those of others) and on constructed examples rather than on naturally occurring discourse (whether spoken or written) as evidence for their theoretical claims.

Lightfoot addresses an interesting question with implications for the understanding of language acquisition and of claims about Universal Grammar (and innateness). I am ultimately unpersuaded.

**2. *"What's in the LAD? – "Another black box."*** Linguists generally subscribe to some version of a view which says: (1) Infants immersed in talk themselves generally acquire language; (2) they do this by processing what they hear through a "language acquisition device" (sometimes disdainfully referred to as "the black box"). Specification of the contents of the black box or LAD is a goal of all linguists of "innatist" persuasions (most contemporary linguists); they seem to agree that they'll find some sort of genetically programmed "software" which permits learners to construct theories of whatever language they hear. It seems to me that Lightfoot is further saying that (3) to the extent that linguists subscribe to such a position *and* to the extent that they agree with the "poverty-of-stimulus" codicil to the theory, they must assign greater magnitude to the genetic contribution to language acquisition. Whether researchers have started from "selectionalist" or "binding" perspectives, moreover, it seems to me that the argument just reviewed has led to the postulation of more and more features which the UG "must" have – in boxes further and further "down" (or "in"?) – and thereby to attempts (1) to specify what data could "trigger" incorporation of those UG features into individuals' competences and (2) to demonstrate that things done in rather different ways in quite disparate languages are actually, for purposes of triggering, the same. The data used in attempts to specify and demonstrate have ordinarily been constructed utterances (or fortuitously overheard ones). To my knowledge the idea that features of UG might result from such extralinguistic attributes of good theories as parsimony or efficiency is not entertained by linguists.

**3. *Some differences in the semantic domain of "obvious."*** Space limitations for *BBS* commentaries do not permit me to detail the reasons for my discomfort about observations/claims that Lightfoot adduces, usually uncritically, in the course of his argument. I'll simply list some of them and provide brief reactions. Lightfoot either explicitly states, or implies, the following:

(1) Constructed examples, based on linguists' "intuitions," constitute adequate bases for the investigation both of features of formal grammar and of processes of language acquisition (*passim*). In my view it has been demonstrated that this is a most dubious assumption (e.g., Bowerman 1985; Brown 1973; 1977; Ervin-Tripp & Mitchell-Kernan 1977a; 1977b; Fillmore 1979; Ochs 1986; Schieffelin & Ochs 1986; Slobin 1985a; 1985b; Snow 1977; Snow & Ferguson 1977).

(2) "Most ambiguities pass unnoticed and people take the most appropriate of the available meanings" (sect.2, para.4). This is an undocumented and probably undemonstrable claim (Grimshaw 1987; Levine 1985).

(3) "Motherese" is the primary data for children's language learning. Snow (1977) herself makes no such claim (see also case studies as disparate in their authors' orienting perspectives as Bloom 1970 and Halliday 1975). Lightfoot's characterization of Snow more generally seems misleading. Snow emphasized that learning is an *interactive* process in which mothers (or other caretakers) and infants mutually accommodate.

(4) Some children are given no explicit instruction; others are corrected by "zealous" parents, and when corrected resist (sect.3). Children are corrected by a range of interlocutors; they do resist; they also accept – and sometimes adapt or explain (Gumperz & Hernandez-Ch. 1971).

(5) Not only are children not instructed, and when corrected resist, but they also disattend (i.e., don't experience as "triggers") what they hear that is not part of the socially most valued local varieties. This remains undemonstrated.

(6) Lightfoot asks why some speakers hear linguistic forms but do not reproduce them; his answer is that syntactic shifts occur in such a way that utterances previously analyzed in a particular way no longer constitute triggers. Although his claim is not that a change occurs in the UG/genetic program (as seems mind-bogglingly to be the case on a first reading), Lightfoot writes as if the richly documented studies of social constraints on language change (e.g., Labov 1972b; 1980; 1986; Weinreich et al. 1968) had never been done. It may well be that triggering phenomena have nothing at all to do with change (although at some point Lightfoot and his colleagues will have to address the question of whether changes in the UG have accompanied other changes in human genetic make-up over the millenia). For reasons similar to my questioning of his explanation for language change, I find Lightfoot's proposals concerning creolization less persuasive than the more socially oriented views (e.g., Sankoff 1980).

At one point in his argument Lightfoot reports a hunch that an explanatory scenario is "too exotic" and that an interpretation based on UG is more persuasive (sect.4, para.9). At a later point, where he reports differences in production among Italian speakers (a fact which counters a strong UG position), he remarks that the data are "slightly exotic" and that variation is hence not surprising (sect.5.4, para.1). It seems to me that this amounts to saying that when data support the UG thesis, that's fine, but when they don't, the problem is with the data rather than a challenge to the UG perspective. This sounds like wanting things both ways.

**4. *In spite of my complaints . . .*** Although I have complained about issues of documentation/demonstration, and particularly about Lightfoot's apparent unwillingness to take any account of social variables, I nonetheless found this an intriguing and suggestive paper. I like both the mode of argument and the systematic marshaling of data from within the paradigm. I wish there could be some rapprochement between formal linguists at the "autonomous core" and those sensitive to social contexts. At the same time, the facts of language are so impressive that it is not surprising that linguists continue to attempt to open up the black boxes they keep finding inside those already opened. Those same facts also make the argument that there is some genetic programming plausible – even if undemonstrated.

## Positive and negative evidence in language acquisition

Jane Grimshaw[a] and Steven Pinker[b]

[a]*Department of Psychology, Linguistics & Cognitive Science Program, Brandeis University, Waltham, MA 02254 and* [b]*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139*

**Electronic mail:** [a]*grimshaw@brandeis.bitnet and* [b]*Steve@psyche.mit.edu*

Lightfoot's interesting target article is part of a growing body of work that has transformed the study of language acquisition during the last decade. Rather than studying children's speech as an end in itself, this research has proposed explicit learning mechanisms by which the child could acquire grammatical structures upon exposure to input data. Any such model of language acquisition has to take into account the nature of the input data: whether children have access to information about what is a sentence in the language (positive evidence), information about which strings of words are not sentences (negative evidence), or both.

Though no one questions that children hear and learn from positive evidence (parents' speech), the question of negative evidence is more complex. Children end up knowing that certain sentences are ungrammatical, yet many theorists, including Lightfoot, take the position that this information is not present in the linguistic world of children. Somehow, children must generalize beyond what they hear, but not so much that they permanently generate ungrammatical sentences by making generalizations that the world can never disconfirm. A major goal of Lightfoot's proposals, and those of other language acquisition theorists (see Baker & McCarthy 1981; Grimshaw 1981; Pinker 1984; 1989; Wexler & Culicover 1980), is to solve those problems. If negative evidence is available, however, these specific proposals would lose much of their motivation. The key questions are: Does negative evidence exist? Is it there in a usable form? Do language learners use it? Do language learners need it?

Brown and Hanlon (1970) examined parents' feedback to their children's questions and negatives, a collection chosen because these constructions are frequently ill-formed in child language. Brown and Hanlon divided adult responses by appropriateness into "sequiturs" and "nonsequiturs" on the assumption that children might take comprehension failure as evidence that their utterance was ungrammatical. Brown and Hanlon found no correlation between the grammaticality of the child sentences and appropriate interpretations by adults. They also studied "contingent approval" to see whether parents tended to respond with disapproval to ungrammatical sentences. Again, they found no relation.

These results are the empirical foundation for the assumption that children do not receive negative evidence. But the conclusion has not sat well with developmental psychologists. Recently, four studies have reopened the question. Hirsch-Pasek et al. (1984) studied approval and its relationship to well-formedness. Here their results accorded with those of Brown and Hanlon. However, they also tallied repetitions (both verbatim and with modifications), and found that adults are more likely to repeat ungrammatical than grammatical sentences. Specifically, they found a statistical relation between maternal repetition and ungrammaticality for their 2-year-olds and none for older children (3–5 years): 20.8% of the ill-formed sentences of 2-year-olds were repeated whereas only 12.0% of the grammatical ones were. Demetras et al. (1986) found that most measures of parental behavior were independent of the grammaticality of the 2-year-old children's sentence, but some were not: Verbatim repetitions and continuations of the conversation followed grammatical sentences slightly more often than ungrammatical ones, but clarification questions followed ungrammatical sentences slightly more often.

Similar results have been documented by Bohannon and Stanowicz (1988) for 2-year-olds, and by Penner (1987) for 2-year-olds, and, greatly attenuated, for 3-year-olds. None of the studies found any form of feedback that followed only grammatical or only ungrammatical sentences; it was only relative frequencies that differed, and usually by only a few percentage points. The conclusion drawn by some of these researchers is that feedback giving information about ungrammaticality is available to learners and that the claims of theorists like Lightfoot are unmotivated.

One thing is certain: Children do not receive negative evidence in the technical sense. Negative evidence (see Gold 1967; Pinker 1979) is information about the ungrammaticality of every ungrammatical string composed of the language's vocabulary items. None of these studies have shown that all of children's ungrammatical sentences elicit differences in parental behavior, only that some do. Furthermore, feedback could only serve as true negative evidence if children could utter any string of words whatsoever and expect appropriate feedback. Note that most of the ungrammatical strings that linguists try to account for are not uttered by children in the 2–3 year age range for which feedback has been documented; there is little reason to believe that children could count on feedback if they uttered them.

An equally important question is whether all parents provide these kinds of feedback. A number of the researchers use statistics to show that the sample data come from maternal populations in which the mean number of responses to grammatical sentences and ungrammatical sentences differ. But these statistics are beside the point: The population means may differ while some of the individual mothers may make the responses equiprobably (or with a difference in the opposite direction). Yet their children, presumably, learn to talk. Unless all mothers of children who acquire language provide feedback, we cannot attribute an important role to it. Not only are individual data not reported in most of the studies, and standard deviations quite large, but caretakers from the socioeconomic classes or cultures that may be less apt to lavish moment-by-moment attention on their children have not been studied at all.

Even for those children whose parents provide feedback, and for those sentences that elicit it, direct information about ungrammaticality is not available. The learner cannot just conclude that a sentence which gets repeated is ungrammatical, or that a sentence which is not repeated is grammatical; the feedback is a small statistical tendency, and not an absolute relationship. For example, any children in the Hirsh-Pasek et al. sample who changed their grammar so as to rule out a repeated utterance would be making their grammar better a fifth of the time but making it worse an eighth of the time, oscillating between incorrect and correct hypotheses.

It might seem that the problem does not arise if we study, for example, corrections, instead of repetitions. After all, unlike repetition, correction presumably occurs only with ungrammatical sentences, so the child could safely assume that if a sentence is corrected it's ungrammatical. There is a problem with this reasoning, however: How does the child know what to count as a correction? Unless corrections are uniquely marked off somehow, the child cannot know whether or not a response constitutes a correction. A correction, if you don't know that it's a correction, is just a partial repetition. Since other partial repetitions occur, children cannot know which are in fact corrections without knowing whether what they said was ungrammatical. But this is the very information that the correction must supply, if it is to provide negative evidence. For example:

ungrammatical: Daddy put the book.
correction: Daddy put the book on the table.
grammatical: Daddy threw the book.
partial repetition: Daddy threw the book on the table.

Calling something a correction in the first place reflects a prior classification of the child's previous utterance as ungrammatical. It's a coding of the data that is available to the experimenter but not to the child. So unless it's possible to show that corrections can be reliably distinguished by children from partial repetitions, the supposed negative evidence would not exist in any form accessible to the learner, since the learner does not in fact know that something is being corrected, and cannot therefore use the information that a correction has occurred.

Perhaps, one might argue, over time children could aggregate information from the statistical tendencies uncovered in the cited work, rejecting a sentence if it had been followed by a given type of feedback so often that the hypothesis that it was ungrammatical was very much more probable. But how this would work is quite unclear, to say the least. Children certainly cannot aggregate information about feedback to tokens of particular sentences; no sentence is used by a child often enough. So if they use feedback at all, they must lump "equivalent" kinds of sentences together for the tallies. How they hypothesize the right equivalence classes and assign sentences correctly to them simply reraises all the questions about generalization that nativist theories seek to answer.

The usefulness of the information that a kind of sentence is ungrammatical is highly questionable too. Sentences are generated by large numbers of rules and principles that vary crosslinguistically, not just by one rule, so even a child able to make a binary good/bad decision faces a formidable example of the "blame-assignment" problem in singling out which rule to change or discard. In fact, the feedback documented by some of the studies does not even distinguish "errors" that are due to syntax from those due to bad pronunciation (e.g. bawoon for balloon), so the child cannot even directly relate an error signal to a deficiency in grammar.

The next question about negative evidence, if it exists in usable form, is whether children use it. Lightfoot reports a typical anecdote showing that overt corrections are fruitless, and no evidence better than such anecdotes exists. Though it is possible that corrections or that other forms of enriched interaction with children may lead to changes in their speech, such evidence would have to be interpreted cautiously. Any correction by its very nature also offers positive evidence, and positive evidence of a peculiarly relevant kind. So any study which purports to show that corrections are actually used by children can be given the more parsimonious explanation that this is another case where relevant positive evidence is used. To make any kind of case for the role of corrections as negative evidence, it is necessary to distinguish the statistical correlation of partial repetitions with ungrammatical utterances from the content of partial repetitions themselves – a form of positive evidence.

Finally, even if children do use feedback, the most germane question is whether they have to. Consider the strings that adults come to recognize as ungrammatical (any of the asterisked example in Lightfoot's paper, for example). Have all speakers of English uttered such strings in their lives and accumulated negative feedback showing that they are ungrammatical? Would an adult who had never uttered the sentence type, or who had not been in the company of people who had noticed and reacted to it, find it grammatical? The relative uniformity of people's judgments of grammaticality for unusual sentence types makes this highly unlikely.

The conclusion, then, is that the newer studies do not provide the slightest reason to question the research program Lightfoot advocates and illustrates, with its assumption of no negative evidence. But it is important to consider how central the absence of negative evidence is to the current research program in learnability. Although individual arguments may rest on the assumption that negative evidence is unavailable, the research program as a whole certainly does not. After all, linguists have access to all the negative evidence there is, but they are still remarkably less good at grammar discovery than children are. The real question concerns the difference between a child and a linguist. Why is the child so much more successful? The answer is surely the one that nativist theories of language acquisition provide – the child has Universal Grammar to use, whereas the linguist is trying simultaneously to construct grammars and to discover the properties of Universal Grammar.

## The language learner: A trigger-happy kid?

Yosef Grodzinsky

*Department of Psychology, Tel Aviv University, Tel Aviv 69978, Israel; Aphasia Research Center, Boston University School of Medicine, Boston Veterans Administration Medical Center, Boston, MA 02130*
**Electronic mail:** *kid@taunds.bitnet*

A principles-and-parameters theory of grammar calls for a triggering theory of language acquisition of the general kind Lightfoot describes. A full triggering theory has to specify not only what the trigger is – to demonstrate that the child can identify his grammar on simple data – but it must also provide the selection device by which the child attends to the relevant data and sifts away the unimportant and ungrammatical. We need, in short, not just convenient conditions for pulling the trigger, a hard enough problem to solve, but also directions for when to withhold. To construct such a theory is extremely difficult.

Lightfoot adduces evidence from historical change to show that some of the child's linguistic experience is not used. This claim is quite plausible, and, as Lightfoot points out, other evidence (for example, the fact that kids ignore ungrammatical strings that appear in the text they are exposed to) points in the same direction. This raises two important questions. First, which portion of the text is sufficient for language learning? Second, given that it constitutes only a subset of the linguistic material the learner hears, how does he identify it and avoid the rest? Lightfoot has an interesting proposal regarding the first question, but not the second. We do not know what prevents the child from assigning a grammatical analysis to "noise," which may lead to wrong conclusions about the grammar of his ambient language. So, whereas the target article attempts to identify the simplest triggering data that would suffice, the issue is not just what the trigger is, but also what isn't the trigger, and how the child knows it. This is a serious problem, given that the text contains ungrammatical utterances, hesitations, and false starts galore, which the child presumably knows he should not misanalyze, but simply ignore. Lightfoot seems to recognize the fact that a trigger-happy language learner would end up excommunicado – with the wrong grammar. Miracles aside, we would like to know how this is avoided.

A concrete example will make my point clear. Consider the well-known *give/donate* contrast that has occupied students of language acquisition for a long time:

(1) a. John gave a book to Bill
    b. John gave Bill a book
(2) a. John donated a book to Bill
    b. *John donated Bill a book

The problem here concerns how the child knows the relation between (1a) and (1b), but blocks (2b) without negative evidence. Several solutions have been proposed in the literature. Of these, Randall's (1985) "retreat route" idea is an appropriate one in the present context. She accounts for the child's "retreat" from overgeneralization in this case by assuming an "Order Principle" that disallows optional complements to intervene between a head and obligatory complements. This principle thus relates the facts in (1) and (2) to the contrast in (3):

(3) a. *John gave a book
    b. John donated a book

Since the NP with the Beneficiary θ-role of the verb *donate* is optional (as (3b) shows), the Order Principle blocks (2b). This is not the case for *give*, however, which obligatorily takes two NPs. Thus the appearance of sentences like (3b) in the child's environment should suffice for him to know that (2b) is blocked, and no negative evidence is necessary.

The descriptive generalization captured by the Order Principle may follow from something else. In fact, it can be seen as a consequence of Larson's (1988) recent analysis of double-objects. This analysis assumes verb-Raising in predicates like *give* (4b), with an underlying structure as in (4a), and a thematic structure as in (5):

(4) a. John [a book [gave to Mary]]
b. John gave [a book [t to Mary]]
(5) give–Beneficiary
Goal of motion along some path

The relevant part of this analysis for us is the claim that the preposition *to* also assigns the θ-role of "goal of motion along some path," and that in *give*, this role is subsumed by the predicate, whereas in *donate* it is not. This is the difference between *give* and *donate* according to Larson. Because of this difference, dative shift in the latter is impossible (2b) because it "violates (the equivalent of) 'recoverability of deletion'" with respect to θ-roles. This claim also explains the contrast in (3), because if the second object of *donate*, unlike that of *give*, depends on the preposition thematically, then its oblique object is always optional, whereas *give* must have an oblique object to discharge all of its θ-roles. Thus (3a) violates the θ-criterion, whereas (3b) does not. On this view, the Order Principle now reduces to differences in θ-grid between the two verbs, and it is this difference that the child has to identify. This identification would presumably come – as Randall proposes – from starting with the assumption that all dative verbs are double objects (like *give*), and then being exposed to examples like (3b), which lead to a revision of the lexical entry.

Yet it is not unreasonable to assume that the child hears ungrammatical expressions like (3a) quite often. If not in isolated form, then certainly in discourse, the rules of which he surely doesn't know yet. Worse yet, he might be exposed to sentences like (6), where there is nothing – no semantic, intonational, or any other cues – to tell him whether (6a) or (6b) is the right analysis:

(6) a. Mommy gave [the baby] [food]
b. Mommy gave [the babyfood]

Why, then, do children not take such misleading utterances into consideration? If Randall is right, and if data such as (3a) do appear in the text, then the child, unaware of the ungrammaticality of (3a), should misanalyze *give* as having an optional complement. This mistake would lead him to treat *give* as if it were *donate* – taking one obligatory complement NP. However, upon facing examples like (1b) in the text, he would be led to the opposite conclusion, namely, that both complements of *give* are obligatory. His lexicon at this point should contain a contradiction. With the absence of additional evidence, he would therefore be forced to reject either (1b) or (3a). Yet there is nothing in the input that would tell the child where to go, which should lead him to an arbitrary choice – concluding, at random, that one of these two sentences is ungrammatical. This does not happen, obviously – English-speaking kids do not converge on different grammars in this respect. They are much more cautious, pulling the trigger on data of the right kind only. Why, then, do all kids behave the same way and choose (1b)?

It must be emphasized that the foregoing is an instance of a much broader problem. Similar considerations hold, for example, of Pinker's (1988) recent proposal that the child has a "learning sequence" that records argument structures of verbs in parental speech, counts them, and creates rules on this evidential basis. In this case the problem of selecting the right data base arises just as forcefully. Whereas the cases discussed deal with lexical entries, one can easily imagine how the same issues arise with respect to parameter fixing. It is quite clear, then, that a solution – a selection device – is necessary for a learning theory to get off the ground.

A good answer is not in sight, of course. One can list the logical possibilities, however. Two kinds of solutions are conceivable for the problem of input selection: "Internal" ones (assuming that the child has an ability [criteria] to select the good portion of the text) and "external" ones (assuming that the bad portion of the text comes somehow marked, in a way the child is capable of identifying).

The external account has three versions, none of which seems particularly promising. The first is based on the claim that the input is presented to the child in an ordered fashion, simple constructions first, complex ones later; the second is based on the effects of the frequency of occurrence of a given expression in the text the child is exposed to; the third is based on the hypothesis that the input comes marked – the child notes ungrammaticality by deviant intonational contours and the like. Arguments of the type Lightfoot gives against the "motherese" hypothesis seem generally valid here too, decreasing the likelihood that the "external" account will bear much fruit.

The "internal" account requires a mechanism by which the child highlights good data and throws away the bad without analyzing it. It is, in fact, a complicated form of "bootstrapping": The child has to decide what the right data are for a grammar he does not know yet. Whereas such a solution is imaginable, we are clearly still far from it at this stage.

The problem is hard, yet a solution must be found, lest we back ourselves into the corner where everything is innate, being forced to a position that even Fodor would not endorse: that not only are concepts like "syllabify" innate, but also their lexical entries – which is, really, a bit too much.

ACKNOWLEDGMENTS

The preparation of this manuscript was supported by NIH grant 11408, by the Israel Institute for Psychobiology under a grant from the Charles Smith Family Foundation, and by the Bat Sheva de Rothschild Fund for the Advancement of Science and Technology.

## Language acquisition: What triggers what?

Hubert Haider

*Department of Linguistics, University of Stuttgart, D-7000 Stuttgart, Federal Republic of Germany*

1. Impressed by David Lightfoot's elegant presentation of the current conception of UG (Universal Grammar) I feel tempted to ask, as an advocatus diaboli, whether it is true that the difference between English and the flu is just the length of the incubation period. By analogy with antibody formation, grammar formation is depicted as the result of a selective process. "The environment may be said to 'select' particular values for the parameters of UG." This analogy fades out the role of cognition, however. UG is a specific system of constraints for symbol manipulation. As a module, it is connected with, and mediated by, less task specific mechanisms of the general cognitive system. UG does not have *direct* access to the primary trigger experience. Chomsky (1981a, p.10) accordingly stresses the "epistemological priority" of parameter values. These must be identifiable from a nonlinguistic categorization of patterns.

A plausible alternative hypothesis would be this: UG is an innate piece of software (an inborn program) for building up a cognitive structure called grammar. This cognitive structure determines the verbal behavior. The analogies used by Lightfoot suggest a view of UG as a hardwired system, parts of which

will be activated upon exposure to a specific input whereas others may atrophy. [See Ebbesson: "Evolution and Ontogeny of Neural Circuits" *BBS* 7(3) 1984.] If this is so, language acquisition should be a continuous and irreversible process that stops when the last parameter is fixed. What we observe, however, is step-by-step development, with frequent reorganizations (cf. Pinker 1984), suggesting that parameters may be fixed with improper values and then reset in subsequent stages. This would fit with a concept of UG as an innate program for grammar formation. UG could be said to guide the child toward projecting specific structures on selected data: the trigger experience. In the course of language acquisition, UG focuses attention on specific data structures. This would account for both the stage specific frequency of patterns and the child's resistance to untimely correction (sect.3). One can sum up this proposal in the form of a brief answer to the initial question: UG triggers *the child* to scan the environment for what is called the trigger experience.

2. How complex are the data that are to be scanned for adjusting the parameters? Lightfoot's position, well substantiated by a series of case studies, is the strongest possible one: No complex data are needed. Any linguist will try hard to concoct some counterarguments to undermine a bold hypothesis. Here are some: Let us start with some general considerations. In any natural language, the subset of sentential structures characterized by degree-0 complexity is finite. It can easily be enumerated because there is no recursion. It is highly plausible that embedded clauses are determined by the same kinds of principles that apply in nonembedded clauses (cf. the principle of cognitive harmony) except for one property, the nonclause bound relations. These can arise only in recursive structures. One might therefore expect that the trigger experience must be sought at least in a subset of degree-1 complexity. Lightfoot suggests an interesting characterization of a degree-0 domain that consists of the main clause plus a bit of the embedded clause. This is the local domain of binding. But even if the strong position cannot not be maintained, Lightfoot's way of defining a local domain is well chosen. It turns Wexler and Culicover's (1980) degree-2-domain immediately into a degree-1-domain, which is a most welcome result. If degree-2 were necessary, one might ask why not any other degree, since the choice of numeral 2 seems arbitrary. Degree-1 simply means a domain with embedding in contrast to a nonembedding domain.

Scandinavian languages seem to abound with nonlocal phenomena. I will briefly illustrate two phenomena whose parametrization seems to violate the degree-0 restriction, namely, long distance reflexives and intersecting extractions. The following examples are taken from Icelandic.

(1) a. Jón$_i$ segir að María elski sig$_i$
John says that Mary loves himself (Anderson 1986, p. 66)
b. Hann$_i$ skipaði Haraldi$_j$ að raka sig$_{i,j}$
He$_i$ ordered Harold$_j$ to shave himself$_{i,j}$ (Maling 1986, p. 61)

In both cases the most local subject accessible for the reflexive is not the matrix subject and hence there is an embedded binding domain. The reflexive may nevertheless be bound, as indicated in (1), by the matrix subject. In order to set the parameter that distinguishes English from Icelandic, it seems that a degree-1 domain must be taken into consideration.

The example in (2), taken from Norwegian (Koch Kristensen 1982), illustrates intersecting extractions. English allows only nested extractions (cf. Pesetzky 1982), and only with nonfinite complements (cf. 3a vs. 3c). In a nested extraction (3a) the path between the highest extraction position and its extraction site includes the other path. In an intersecting extraction (3b) the paths do not overlap completely.

(2) Hvilke malerier$_i$ har ikke Petter noen vegg$_j$ å henge opp e$_i$ på e$_j$
which paintings has not Peter any wall to hang up on

(3) a. Han hade insett på ett tidigt stadium att han maste forstora banden.
He had realized at an early stage that he must destroy tapes.
b. Nixon sade att han redan på ett tidigt stadium (hade)
Nixon said that he already at an early stage (had)
insett att han maste forstora banden.
realized that he must destory tapes.

The Norwegian example indicates that the original ban against intersecting extractions (cf. Pesetzky's path containment condition) must allow for parametrization. In order to fix this parameter for Norwegian, it is an embedded binding domain that is again crucial. It seems that the proper relation between the extracted phrases and their respective extraction sites can be determined only if a degree-1 domain is accessible.

Even if it should turn out that these cases resist a degree-0 treatment, Lightfoot's proposal of the relevant locality domain together with a degree-1 domain for the trigger experience is a reasonable result. I wonder how a degree-2 phenomenon might look. It should be beyond comprehension for a UG-driven system.

## Degree-0 explanation

Roy Harris

*Department of English Studies & Comparative Literature, University of Hong Kong, Hong Kong*

Inviting a sceptic to comment on the internal niceties of generativist language-acquisition theory is rather like asking an atheist to discuss rival theologians' various interpretations of the doctrine of transubstantiation. If no such process ever takes place, then all accounts of exactly *how* it takes place must be mistaken. Conceivably, certain orthodox accounts are more mistaken than other orthodox accounts; but the interest in degrees of compound error is marginal. From a radical unbeliever's point of view, the only interesting question is why anyone should believe the orthodox doctrine in the first place.

The answer is not far to seek in the present instance. As Lightfoot's introduction amply demonstrates, some linguists have fallen for – and deliberately revamped – a metaphor familiar to biologists. Anyone acquainted with the history of nineteenth- and twentieth-century linguistics will understand why. It is part and parcel of the long drawn-out academic battle to obtain for linguistics recognition as one of the natural sciences. The strategy is to show that the methods and assumptions of linguistics parallel those of the established scientific disciplines.

The key phrase "triggering experience" arguably makes some sense if we are discussing how fledgling birds learn the birdsong patterns of their species [see Johnson: "Developmental Explanation and the Ontogeny of Birdsong" *BBS* 11(4) 1988.]; but it makes no *prima facie* sense at all in the case of a human child learning the syntactic details of English, French, Cantonese, and so forth. For one thing, these languages are not species-specific. For another, every parent knows that a child's first steps across its family's linguistic threshold are taken when the child's voluntary (but encouraged) attempts to imitate certain single arbitrary vocables become acceptably successful. If Nature had really been generous enough to provide *Homo sapiens* with built-in aids to language acquisition, it is here that one would expect to find clear evidence for them; whereas knowing how to construct subordinate clauses (which is Lightfoot's main concern) is about as likely to be genetically built-in as knowing how to tie a reef-knot.

Lightfoot clearly has not taken on board any of the basic criticisms which can be – and have been – made of his chosen approach to language acquisition. (Romaine 1985 provides a lucid summary.) For Lightfoot nothing essential seems to have changed since the publication of Hornstein and Lightfoot

(1981). However, the "problem" of language acquisition, as posed within this theoretical framework, has increasingly come to be seen as irrelevant to any enterprise except the obsessive self-justification of generativist linguistics.

Although Lightfoot refers to the child as a "little linguist" in scare quotes, presumably in order to distance himself from the original Chomskyan thesis about language acquisition, it is clear that only "little linguists" can benefit from the putative "triggering experiences" which Lightfoot postulates. For those experiences, if they are to be crucial, have to be subject to an analysis which parallels suspiciously closely the overt arguments which linguists like Lightfoot come up with. The difference between a child "inferring" the rules and having them genetically "triggered" thus turns out to be merely terminological.

In any case, the "triggering experience" hypothesis is explanatorily vacuous in the case of subordinate clauses. For implicitly presupposed are the following assumptions: (i) that any incipient grammar is already "a finite algebraic system," (ii) that what the child hears are utterances with a determinate semantic structure, and (iii) that the child has prior access to a cognitive system of remarkable complexity, which enables unknown factors in the situational context to "force" a preferred interpretation of the structural semantics. But if Santa Claus already brings this much for free, he does not need to include in the genetic Christmas stocking a gift package to help all God's children with the acquisition of subordinate syntax. Or if he did, he might need Occam's cordless shaver as a useful Christmas present himself.

The trouble with the entire generativist scenario of language acquisition is twofold, and the twin faults are perspicuously exemplified in Lightfoot's account of generativist reasoning. (i) The evidence adduced is a priori, even when claiming to be empirical; and the a priori material on inspection reveals closet prescriptivism. For example, Lightfoot finds no fault with the Chomskyite claim (Rizzi 1982b) that one difference between acquiring English syntax and acquiring Italian syntax might be explicable by the fact that Italian children hear sentence-types such as *ho trovato il libro*, whereas English children do not hear **found the book*. The choice of this particular construction is revealing. According to prescriptive grammarians *ho trovato il libro* is "good" Italian, but **found the book* is not "good" English: i.e. it is "ungrammatical" in English to omit the subject pronoun. (How many non-English-speaking linguists have been taken in by this notorious example it would be difficult to estimate.) The fact is that *found the book* (with no subject pronoun) exemplifies a combinatorial pattern frequently encountered in colloquial English (and hence, presumably, heard by English-learning children), particularly in response to questions. Why any reputable theorist should asterisk such expressions is a prescriptivist mystery. But such idealizations are apparently essential to explain generativist "parameter setting." (ii) The rationale on offer is boot-strappingly tautological. For instance, discussing the blackboard example *an old man from the city* (Scott Fitzgerald or the *Financial Times*?), Lightfoot apparently endorses the claim that it can be "generated" by either of two "sets of rules." He writes: "It can be shown that any noun phrase that occurs in English, and thus any noun phrase that an English-speaking child will hear, can be generated by both sets of rules." The give-away word here is *thus*. In short, children who acquire English noun phrases are deemed to have been (relevantly) exposed to (not all but) only English noun phrases as possible "triggering experiences." Anything else the child hears, it will already recognize, evidently, as not an English noun phrase. *Mutatis mutandis* the same presumably goes for any piece of English syntax. Given this assumption, one may reasonably ask whether children need Santa Claus's subordinate-clause "degree-0" gift kit at all. Like most unwanted Christmas presents, it is likely to remain neglected in a corner until someone has the sense to throw it out, along with other degree-0 explanations in linguistics.

## A possible mathematical specification of "degree-0" or "degree-0 plus a little" learnability

Aravind K. Joshi

*Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA 19104*
**Electronic mail:** *joshi@linc.cis.upenn.edu*

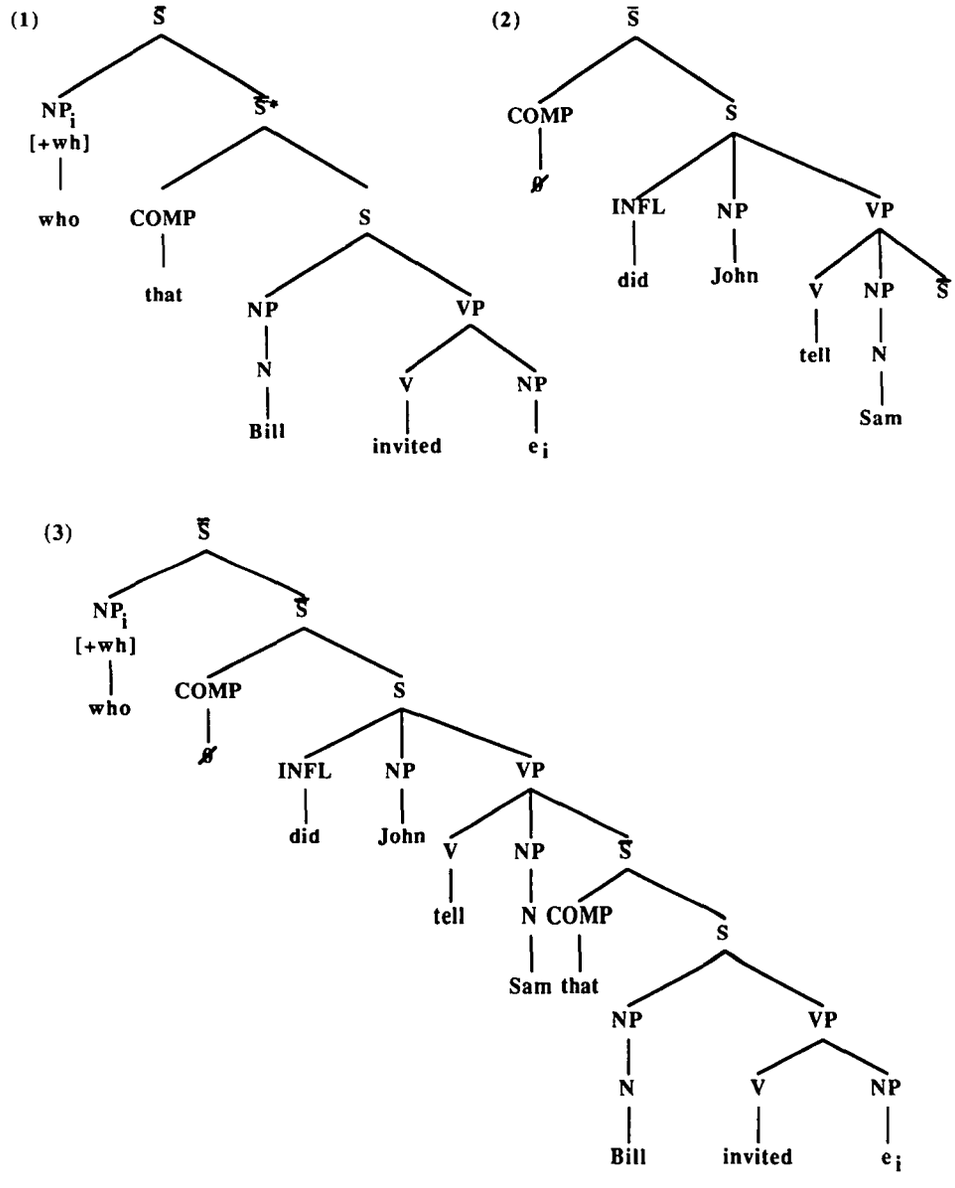
The basic idea discussed in the target article is that all information needed to acquire a grammar can be obtained by the learner from single unembedded structures. Information obtained from these simple structures, together with some general principles that are part of a Universal Grammar (UG), is all that is needed for grammar acquisition. Let us try to see what kind of formal system(s) can support such a learning paradigm.

Let us consider a formal system (formal grammar) which consists of a finite set of (elementary) structures and a single operation of composition of these elementary structures and structures derived from them. If this single operation of composition is considered as belonging to Universal Grammar then the finite set of elementary structures completely specifies the grammar. What is needed to specify these elementary structures? First, of course, we need information about the shapes of the possible elementary structures and constraints that must hold among the nodes of a given structure. We may also need to specify for each node of each structure which other elementary structures are combinable at that node (if there are constraints on combinability). Given such a system (of course, assuming that it is linguistically adequate), it is of interest to see what kind of mathematical system would serve as a specification for the model proposed by Lightfoot.

One such formal system is tree adjoining grammar (*tag*), which has been investigated in detail in recent years; see, for example, Joshi (1985; 1987), Joshi et al. (in press), Kroch (1987), and Kroch & Joshi (in press), among others. A *tag*, G, consists of a finite set of initial structures (trees), I, and a finite set of auxiliary trees, A. The initial trees roughly correspond to minimal sentences and the auxiliary trees roughly correspond to minimal recursive structures. An initial tree is rooted in S or $\bar{S}$ and the frontier is all terminals. An auxiliary tree is rooted in some nonterminal and the frontier is all terminals, except one symbol, which is a nonterminal and the same as the one labeling the root of the auxiliary tree. The trees in I and A together constitute the finite set of elementary trees. There is a single operation of composition, called "adjoining," which adjoins an auxiliary tree to a node in an elementary tree or a tree derived from it. Adjoining inserts the auxiliary tree at a node (with the same label as the label of the root node of the auxiliary tree) in a given tree. Thus, for example (see Kroch 1987), the auxiliary tree (2) when adjoined to an initial tree (1) at the node interior node S (designated by *, for convenience), results in the derived tree (3). One of the key properties of *tag* is that it factors recursion from dependencies (such as subcategorization, agreement, and filler-gap), thus making all unbounded dependencies local in a sense. It is this property that plays a key role in our further discussion.

***Subjacency as a constraint on simple unembedded structures.*** As shown above, wh-movement can be accomplished in a *tag* by the introduction of auxiliary trees. The unbounded character of wh-movement constructions will then follow from the nature of the adjunction operation whereby matrix predicates can be adjoined "between" the wh- in COMP and the embedded sentence. One interesting consequence of expressing long-distance wh-dependencies in this way is that some of the island constraints on extraction, especially Subjacency, are statable as constraints on elementary trees (see Kroch 1987). Consider, for example, a wh-island violation like the following:

(4) *Who(i) did you wonder why she wrote to e(i)?




**Who did John tell Sam that Bill invited**

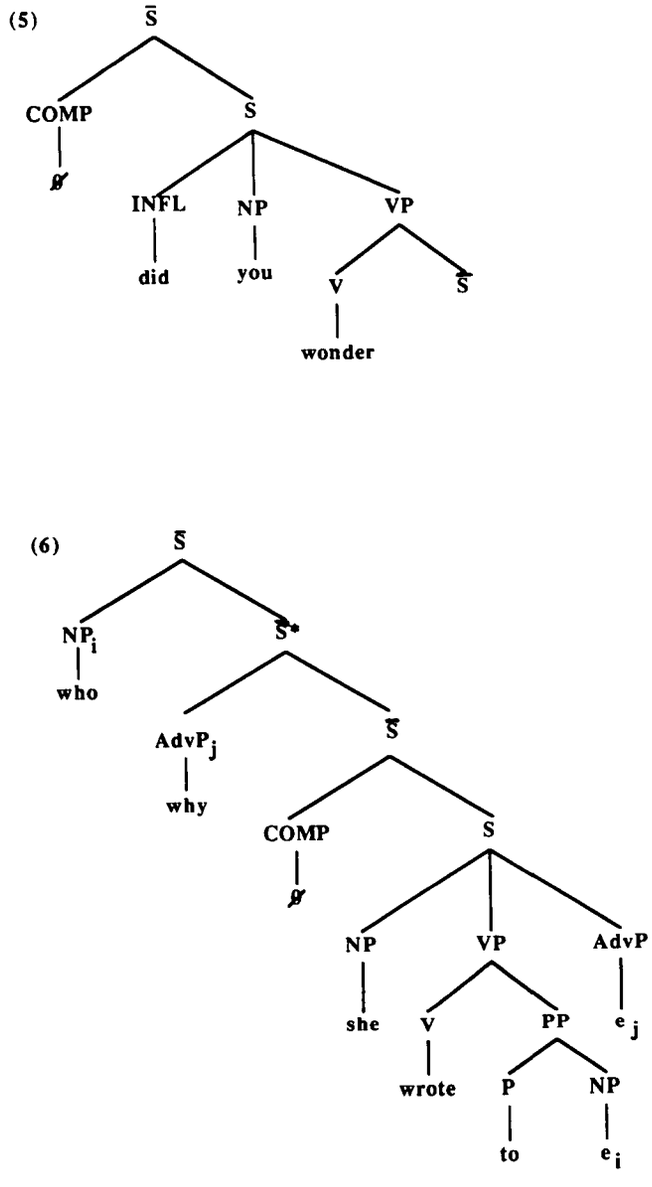
Figures 1, 2, and 3 (Joshi). A derivation in a *tag*.

This sentence cannot be produced by a *tag*. Since long-distance wh-movement is always the result of adjunction of a matrix predicate above an embedded clause, (4) could only be produced by the adjunction of the auxiliary tree (5) to the initial tree (6) at the starred node S. Such a derivation would be illicit since the initial tree here has two preposed wh-phrases, which English sentences never allow. Since the constraints on the well-formedness of the initial trees will have to rule out such configurations to correctly generate simple sentence questions, the ungrammaticality of wh-island violations follows directly.

The significance of this observation of Kroch in the context of the target article is as follows. The learner does not need evidence from complex sentences to induce the Subjacency constraint. Since the learner will have evidence from unembedded structures with only one wh-phrase and will not see structures with two wh-phrases (for English), information about the Subjacency constraint has been implicitly presented to the learner via trigger experiences confined only to unembedded structures (elementary trees of *tag*). Kroch (1987) has shown that several other constraints – for example, ECP (empty category principle) and CED (condition on extraction domain) – can be reduced to constraints on initial structures.

In the example above, tree (5) is adjoinable at the starred node S- in the tree (6) and not to the root node of tree (6). Thus in the specification of tree (6) we must have this information specified in some form, either explicitly as described above, or implicitly by a specification of features at the starred node such that their compatability or noncompatability with the features specified at the root nodes of auxiliary trees will have the effect of stating this constraint. In this sense, in the language of the target article "degree-0" or "degree-0 plus something else," specification of the unembedded structures is needed.

The *tag* system described above is a very highly (mathematically) constrained system. It is more powerful than context-free grammars but only slightly so, and belongs to the class of formalisms known as mildly context-sensitive grammar formalisms. The properties described above follow from the way *tags* factor recursion and dependencies and the domain of locality the elementary *tag* trees provide. This domain of locality is larger than that provided by context-free grammars. This




Figures 5 and 6 (Joshi). Subjacency in *tag*.

extended domain of locality and the factoring of recursion and dependencies is what makes it possible to state constraints such as Subjacency on the elementary trees (unembedded structures). This is exactly what Lightfoot's model requires.

In addition to the mathematical properties described above, it has also been shown recently that *tags* belong to the class of grammars known as polynomially learnable. Polynomial learnability is a complexity theoretic notion of feasible learnability (see Valiant 1984, for example). Abe (1988), in applying this notion to the evaluation of grammatical formalisms for linguistic description, has shown that a novel nontrivial constraint on the degree of "locality" of grammars allows *tags* and some related, mildly context-sensitive grammar formalisms to be polynomially learnable.

In summary, I have presented a formal system which can be regarded as a (mathematical) specification of the model presented in the target article. Since the model of "degree-0 or degree-0 plus a little" is not mathematically precise (and it may be very difficult to give such a formalization), it is not possible to show an exact correspondence between the *tag* formal system and the model of the target article. However, I would not be surprised if it turns out that to the extent the "degree-0" or "degree-0 plus a little" model can be made mathematically precise, it will correspond to grammars that factor recursion and dependencies and require a domain of locality which is larger than that corresponding to context-free grammars.

## Does Universal Grammar exist?

Jan Koster

*Department of General Linguistics, University of Groningen, 9712 TG Groningen, The Netherlands*
**Electronic mail:** *juliet19@hyrrug5.bitnet*

Degree-0 learnability of natural languages is an important notion because it means that languages could be learnable on the basis of very simple data, particularly data without embedded clause structures. Lightfoot convincingly shows that the hypothesis of 0-learnability can be maintained if we define it in terms of minimal binding domains rather than in terms of single clauses. I am not aware of convincing counterexamples against the hypothesis of 0-learnability. So far, then, I fully agree with the author.

Unfortunately, the examples used to illustrate the hypothesis are not all that convincing. There is, for instance, no reason to assume that Italian and English differ with respect to an S/S′-parameter for bounding purposes (Lightfoot's section 5.1). The present consensus is that Rizzi's (1982a) original article overstated the differences between Italian and English. The original idea was that Italian allows violations of the wh-island conditions (such as Lightfoot's example 21.a), whereas English does not. This conclusion appeared to be far too simple. Thus, English allows many relatively acceptable violations of the wh-island constraint as well (as was already pointed out by Reinhart 1975):

(1) a. What don't you know when to file?
b. What don't you know how long to boil?
c. What don't you know where to put?

Reinhart further pointed out that such violations are even possible with tensed embedded questions:

(2) What books don't you remember who borrowed from you?

Many examples of this kind could be added, and it is clear that Italian and English cannot be distinguished in terms of a simple S/S′-parameter.

Similarly, French examples such as Lightfoot's (24) do not necessarily show that French has less strict bounding behavior than English. One problem is that similar examples can be constructed for Dutch or German, which are even stricter than English with respect to bounding. Consider, for instance, the following Dutch examples:

(3) a. [$_{NP}$ Wat voor een boek$_i$ [$_S$ heeft hij t$_i$ gelezen]
what for a book has he read
"What kind of book did he read?"
b. [Wat]$_j$ [$_S$ heeft hij [$_{NP}$ t$_j$ voor een boek] gelezen]

As in French, it is possible in such Dutch constructions to extract only part of an NP (as in 3.b). As in French, one might then conclude that S is not a bounding node in Dutch. In reality, however, the island behavior of Dutch is stricter than what we find in either English or French. So it is not clear what the examples at issue establish. But again, the whole idea of S/S′-parametrization seems to be a nonissue according to our current knowledge.

Another point of disagreement with Lightfoot is the persistent use of the term Universal Grammar (UG) as a characterization of the initial state of the language learner. Although such usage is common practice, it is far from clear whether it makes sense. The term Universal Grammar is problematic

because of the highly modular nature of language. It is generally assumed nowadays that language is what the structuralists used to call a "system of systems." What we call language is the result of the combination of highly diverse and autonomous subsystems. It would be highly exaggerated to claim that generative grammar is in actual practice equally concerned with all subsystems that together constitute natural language.

In actual practice, generative grammar is mainly concerned with the subsystem often referred to as the computational system. To a large extent, this system abstracts away from the conceptual systems provided by the human mind (for instance, those systems that define knowledge of the world). The computational system studied by generative grammarians only describes recursively defined hierarchical patterns of discrete units (such as phonemes and morphemes) and certain relations among these hierarchical patterns. What we call natural language only comes about by combining this computational system with the conceptual systems. It is highly questionable whether one should refer to any of the systems in isolation as language or grammar.

In speaking about Universal Grammar (or even the language faculty) if one actually only speaks about the computational faculty, one seems to take it for granted that the computational faculty has an inherent purpose, namely, its combination with the conceptual systems into what we call natural language in everyday life. In our present state of knowledge it is sheer dogma to assume that the various subsystems of natural language are connected by principles of biological necessity.

Alternatively, it is entirely possible that the subsystems were only combined as an accidental historical step, by human ingenuity. Since, according to this alternative, the basis for the subsystems is still innate, poverty-of-the-stimulus arguments cannot decide the issue. Similarly, the fact that the computational faculty has a very specific application in language does not show that it cannot have other applications. As a matter of fact, the human capacity to count (also based on recursiveness) seems to indicate that certain aspects of the computational faculty can be applied in other domains.

But if the computational faculty is not demonstrably specific to language (as a matter of biological necessity) it does not make much sense to refer to its initial state as Universal Grammar (as Lightfoot and most generative grammarians do). Even if all structural components of language have an innate basis, it does not follow that the act of combining the components into language has an innate basis. Strictly speaking, therefore, it has not yet been established that there is an innate language faculty.

## Language learning and language change

Anthony Kroch

*Department of Linguistics, University of Pennsylvania, Philadelphia, PA 19104-6305*

**Electronic mail:** *kroch@cis.upenn.edu*

Lightfoot's target article points out that there is an inverse relationship between the richness of the child's genetic linguistic endowment, so-called Universal Grammar (UG), and the complexity of the linguistic experience needed by the child to trigger successful language acquisition. Given the generative argument that children must bring to the task of first language acquisition an extremely rich genetic endowment, Lightfoot's observation leads him to propose that the only language specific information needed – or used – in acquisition is present in a severely restricted subset of the forms that language learners are exposed to, namely, in unembedded domains. This hypothesis is certainly plausible. Indeed, it is even expected in light of recent work on the mathematicization of linguistic theory (see commentary by Joshi). On the other hand, Lightfoot's attempt to provide specific empirical support for his hypothesis suffers, perhaps necessarily so, from a reliance on specific details of the form of UG about which generative grammarians disagree, even within his own school of government-binding theory. Therefore, rather than pick apart the target article's specific analyses, which could be modified in countless ways either to preserve or to refute the central hypothesis, I will respond in this commentary to a particularly interesting remark that Lightfoot makes concerning the interaction of language acquisition with historical change, an area to whose investigation he has made a major contribution. I will indicate how certain recent research on syntactic and morphological change supports his general characterization of the language acquisition problem and also his contention that problems of language change are intimately related to problems of acquisition. At the same time we will see that the work suggests a view of the relationship between the two that is somewhat different from the one Lightfoot himself adopts.

In motivating the notion that the data active in triggering acquisition are only a subset of those to which the child is exposed, Lightfoot asserts that the very existence of linguistic change requires that children ignore some of the data in their linguistic environment in the course of language learning. When the language of one generation of speakers differs from that of the preceding generation, at least in simple and commonly occurring contexts, it must be, Lightfoot asserts, that learners have ignored some of the input to which they were exposed. But although there may be some cases that proceed in this way, my own reading and research in historical change leads me to think that change is more often gradual than abrupt and that one generation is more likely to differ from its predecessor in the frequency with which its speakers use certain forms than in whether those forms are possible at all. Only when the frequency of a form drops below a minimum threshold do learners reanalyze their grammatical systems so as to exclude it (Fontaine 1985; Kroch et al. 1982; Kroch 1989; Noble 1985; Oliveira e Silva 1982). Even in cases where a grammar is reanalyzed in the course of linguistic change, it is more than common for features of the old grammatical system to persist as remnants. For example, English, in the course of its history, lost the so-called verb-second constraint that characterizes the Germanic language family; but in a few cases the subject–verb inversion triggered by the constraint still occurs, as in verb phrase deletion examples like (1):

(1) Abigail likes ice cream and *so do I*.

Examples like this suggest that, far from ignoring data that may be incompatible with a new grammatical system, learners are likely to complicate their grammars with rules of marginal coverage simply to preserve details of a disappearing linguistic system to which they have only minimal exposure.

However, neither the gradualness of linguistic change nor the ability of learners to preserve remnants of an archaic grammar threatens the generative picture of language acquisition. Indeed, recent empirical work on the time course of syntactic and morphological changes supports the hypothesis of a highly structured UG. We and others have found that when a language changes gradually, the locus of the change is not the specific linguistic context where the change is manifest but rather the whole grammatical subsystem responsible for the linguistic form in question. To illustrate this result with a single example, we found, in a recent reanalysis of Eleegård's (1953) pioneering quantitative study of the rise of periphrastic "do" in Early Modern English negatives and questions, that the use of "do" as a tense carrier (as measured by the logistic transform of the frequency) increased at the same rate in all contexts. Furthermore, we found that a superficially unrelated change, a shift in the placement of the light adverb "never," proceeded at the same rate as the rise of periphrastic "do" (see Kroch 1989 for details). Under current grammatical analyses (for example,

Roberts 1985) these two changes are reflexes of a single modification in the way that English verbs are inflected for tense and agreement. While Middle English exhibited the process of "verb raising to INFL" (also found in such languages as French), Modern English has lost it and relies on a rule of affix hopping. The loss of verb raising forces both the use of a periphrastic auxiliary as a support for tense inflection in negatives and questions and a change in the position of light adverbs, though limitations of space prevent us from demonstrating the point here. Fontaine (1985) presents a similar case of superficially disparate reflexes of a single grammatical subsystem changing together in her quantitative investigation of the simultaneous loss of subject pronoun deletion and subject postposing between Old and Modern French.

If we ask ourselves why the various contexts of a linguistic alternation should, as a general rule, be constrained to change in lock step, the only apparent answer consistent with the facts of the matter is that speakers learning a language in the course of a gradual change learn two sets of well-formedness principles for certain grammatical subsystems and that over historic time pressures associated with usage (presumably processing or discourse function based) drive out one of the alternatives. We must then ask, however, what would cause learners of a superficially homogeneous dialect to postulate two mutually exclusive grammars for it rather than one grammar which allowed for variation, especially as stable inherent variation is widespread in grammar (Labov 1972a; 1980). This same question, of course, arises with respect to bilingual and diglossic environments, since children in such environments learn the dialects they hear without mixing the rules. In these cases, the phonological distinctness of the varieties may seem enough of a cue to prevent confusion, but it should be evident upon reflection that the appeal to phonology just poses the same learning question at a different level of language structure. In any case, where competing syntactic and morphological subsystems coexist without overt cues, the only answer to the learning question that is currently available is that the learners' innate dispositions, as specified by UG, force them to analyze the competing variants in the linguistic environment as evidence for two linguistic systems.

Thus, although the fact that linguistic change is often gradual does imply that speakers exposed to two grammatical options learn both rather than abruptly abandoning one for the other – which challenges Lightfoot's proposal as to how language change and language acquisition interact – an empirical analysis of how gradual change over time is constrained reveals that the central tenet of modern generative linguistics, the need for a highly specified Universal Grammar, is as crucial to the explanation of language change as it is to synchronic explanation.

## The nature of triggering data

Howard Lasnik

*Department of Linguistics, University of Connecticut, Storrs, CT 06268*
**Electronic mail:** *lasnik@uconnvm.bitnet*

Lightfoot presents important discussion bearing on the paired questions: What does a child learning a language actually have to learn, and what evidence available to the child drives this learning? With respect to the second question, Lightfoot suggests that no negative data, in any form, are available; he further argues that learning can proceed entirely on the basis of "degree-0" evidence, that is, positive instances with no embedding. In making his case, there is one particularly intriguing phenomenon that Lightfoot calls upon without actually analyzing. The phenomenon involves contraction, as in Lightfoot's example (8a), presented here as (1):

(1) *Jay's taller than Kay's (cf. . . . than Kay is)

Lightfoot claims, plausibly, that children are not provided explicit correction that would serve to inform them of the ungrammaticality of contracting the second underlying *is* in (1). He then argues that this information would not be available to children "indirectly" either, where "indirect negative data" arise by virtue of the recognition that an expected form is not exemplified in the input data. Lightfoot indicates that "it is hardly plausible to say that [children] have indirect access to [1]." Based on a discussion in Lasnik (in press), which draws from Epstein (1982), I would like to explore the question that Lightfoot has left us with: How *do* children come to possess the knowledge instantiated in such examples as (1)?

Before examining how this knowledge is attained, it is necessary to specify just what the knowledge is. The crucial property of (1) is that it seems to indicate that contraction of *is* onto a preceding word is inhibited if material following *is* has been removed or displaced. There are several accounts of this property in the literature, but among the most attractive is that of Bresnan (1971). Bresnan observes that the contraction of *is* contrasts with the contraction of *want* and *to* produce *wanna*, as in (2).

(2) You want to see it → You wanna see it

In particular, whereas the former is inhibited by the removal or displacement of material following *is*, as seen above, the latter is inhibited by the removal or displacement of material preceding *to*:

(3) a. You want who to see it
b. Who do you want to see it
c. *Who do you wanna see it

Conversely, removal of material following *to* does not prevent contraction:

(4) a. John thinks I should see it, but I don't want to (see it)
b. John thinks I should see it, but I don't wanna (see it)

Suppose that *wanna* contraction is just what it appears to be: amalgamation of *want* and *to* when they are contiguous. Also assume, with Lightfoot and much of the literature on the topic, that movement leaves a trace. Then (3c) is correctly prevented, since the trace of *who* intervenes between *want* and *to*. (4b) is allowed since there is no such intervening trace. But what of (1) or (5), where there is no intervening trace?

(5) a. I wonder where John is *t*
b. *I wonder where John's *t*

Bresnan argues that with *is* contraction, things are not as they appear to be orthographically. Rather, contraction of *is* is not to the preceding word but to the following one. Now, if, as seems morphologically plausible, contraction onto an empty category is proscribed, all of the ill-formed instances of *is* contraction are accounted for, as is the otherwise surprising minimal pair (4b) vs. (5b). Finally, there is a straightforward explanation for the possibility of *is* contraction in a configuration abstractly like the one in (3):

(6) a. You think who is here
b. Who do you think *t* is here
c. Who do you think's here

Under Bresnan's analysis, the trace does not inhibit contraction because it does not, in fact, intervene between the involved items.

Thus, as Epstein notes, English uses two options made available by Universal Grammar: encliticization (*to* attaching to the end of *want*); and procliticization (*is* attaching to the beginning of the following item). But what ensures that the child learning English will associate the correct type of cliticization with each phenomenon? For example, suppose the learner mistakenly takes *is* contraction to be an instance of encliticiza-

tion. Then the ill-formed (1) and (5b) will mistakenly be assumed to be well-formed, a problematic state of affairs, given that children would receive no direct evidence that the hypothesis is in error. I would like to suggest that one possibility is indirect negative data. Of course, for the suggestion to carry any weight, there must be some characterization of the circumstances under which a construction would be expected by the child, as Lightfoot points out. As noted earlier, Lightfoot suggests that whatever the characterization, it couldn't possibly extend to the phenomenon under consideration here. But suppose that the child expects contraction wherever it is possible. Surely, normal colloquial adult speech is consistent with such an expectation. I suspect that there will be dozens, or even hundreds, of occurrences of (7a) for every one of (7b) in conversation.

(7) a. He's tall
b. He is tall

When presented with (5a), then, the child might conclude that (5b) is not possible, since, if it were possible, it would have been used. All else being equal, this would exclude the encliticization analysis and would thereby motivate the procliticization analysis. Why contraction is preferred where possible might reflect a preference for "reduced" forms more generally. Chomsky's (1981a) "Avoid Pronoun Principle," which favors null pronominals over overt ones, could be seen as a reflection of the same tendency, as could the use of clitic pronouns rather than full forms, except where emphasis or contrast is intended, in languages that have both pronominal types.

An alternative learning procedure is also potentially available to the child. Consider again (6c), which contrasted with the ill-formed (3c). Assuming that the mechanism blocking the latter example is known in advance by the child (i.e., reflects a principle of Universal Grammar), then upon presentation of (6c) the child will have to conclude that the cliticization process at work here is not encliticization, hence must be procliticization. The well-formedness of (6c) thus provides the child with evidence for an analysis entailing the ill-formedness of (1) and (5b). Under this alternative account, no negative data, either direct or indirect, are invoked. (6c), a positive instance, is the triggering experience. From the point of view of Lightfoot's proposals, however, (6c) raises another issue. It does not seem to be a "degree-0" datum. A Domain can potentially include a clause and a portion of another clause embedded in the first. That portion may extend to Comp, and, under certain circumstances, to the embedded subject, but may extend no further. In (6c), however, the crucial informative part of the structure extends at least to the INFL/AGR element of the embedded clause. This suggests that either the degree-0 requirement is too restrictive or that the structure relevant to determining degree is too narrowly defined. There is some reason to think that the latter might be correct. Lightfoot observes that there are languages (such as English) where the antecedent of an anaphor is limited to the minimal finite clause including the anaphor. In other languages, binding can reach as far as the subject of the next finite clause, as in Lightfoot's Chinese example (35). Lightfoot's theory, however, does not seem to allow for a third possibility that is instantiated in a number of languages, where even the *object* of a clause can be bound by an antecedent outside that clause. In fact, in such languages, the antecedent can be indefinitely higher than the anaphor. For example, in certain Chinese dialects, (8) is fully acceptable.

(8) Zhangsan$_i$ shuo $_S$[Mali xihuan ziji$_i$]
Zhangsan say Mary like self
"Zhangsan said that Mary likes himself"

Thus, binding domains can, parametrically, extend far into an embedded clause. (Alternatively, the parametric difference involves not the domain but the *type* of anaphor. The learnability question would be essentially unchanged.) In addition, the selection between the second and third type of binding domain clearly demands data at least as complex as (8). Perhaps "Domain" could be redefined so as to allow such an extension. Then (8) would still count as a degree-0 datum by Lightfoot's criteria. But this moves us quite a distance from the Wexler–Culicover concept of degree of embedding based on the plausible primitive notion "clause." For Wexler and Culicover (1980), (8) would, obviously, constitute a paradigmatic degree-*1* configuration. Now imagine we were to find some phenomenon that crucially involved still further embedding. Once again, we could imagine redefining "Domain" so as to encompass the necessary structure. Or we might introduce a new structural notion, "Realm," that extends two clauses down, and then we could define degree-0 learnability in terms of Realms. This suggests the possibility that Lightfoot's degree-0 proposal might reduce to a question of terminology.

## The true nature of the linguistic trigger

Marjorie Perlman Lorch
*Applied Linguistics, Birkbeck College, University of London, London WC1H OPD, England; Speech Therapy, National Hospital for Nervous Diseases, London WCIN 3BG, England*
**Electronic mail:** *ublvo30@cu.bbk.ac.uk*

Although his abstract, introduction, and conclusion are tantalizingly biological in their presentation of the learnability problem, the details of Lightfoot's argument are worked out in strictly linguistic terms. The target article fulfils the prophesy made by Smith and Miller more than twenty years ago:

> It is somewhat paradoxical to speak of language universals as constituting subject matter for psychology rather than for descriptive linguistics because it is perfectly obvious that language universals will be discovered and substantiated only as the result of the most careful and painstaking linguistic research. (Smith & Miller 1966, p. 6)

Lightfoot has done an admirable job of keeping up the linguists' side of the bargain. But, in attempting a linguistic *characterization* (i.e. grammar) which has learnability as its goal, the problem becomes one of linguistic *realization*. He has convincingly argued for the possibility of degree-0 learnability. The crucial question he asks is "why should children need to hear more than a single clause (plus the front end of a lower clause) in order to hear the effects of all possible grammatical processes in their language?" The question in my mind revolves around the use of the terms "hear" and "clause." Can we transform this theoretical linguistic question into a psychological and neurological question regarding stimulus sensitivity and stimulus strength which may "tell us more about the true nature of the trigger?"

The hope of finding answers to these questions lies in the transformation of the learnability issue from one of linguistic theory based on a nineteenth-century genetic metaphor, to a detailed model of the epigenetic process of maturation in the brain's language acquisition system which is consistent with current notions of biological growth. [See also Johston: "Developmental Explanation" *BBS* 11(4) 1988.]

## INFL', Spec, and other fabulous beasts

James D. McCawley
*Department of Linguistics, University of Chicago, 1010 E. 59th St., Chicago, IL 60637*

Lightfoot takes up several special cases of the problem: What experience does it take for a child to acquire the syntax of such-and-such class of sentences? In each case, Lightfoot's solution posits genetically determined aspects of linguistic structure that

would allow a child to fill in the remaining details of structure on the basis of simple and readily accessible types of linguistic experience. For example, Lightfoot deals with the problem of the constitutent structure of English NPs by taking the skeletal phrase structure rules (example 6) to be genetically determined, leaving the child with the fairly trivial task of observing the word order that fluent speakers use in the various instantiations of those rules. Or at least the task is trivial once the child has accomplished the much less trivial one of identifying relevant fragments of what he hears as instantiations of those rules. To identify the expression *drink milk* as an instance of (6b) with *drink* as its head, the child would have to identify *milk* as not merely a word but at the same time a "maximal projection" of that word, and not identify *drink* as a maximal projection, lest he incorrectly identify *milk* as the head of the expression.

If the child at this point in his development correctly understands the words *milk* and *drink*, he can identify *drink* as the head of the phrase on semantic grounds, namely, that it is a predicate of which *milk* here functions as argument. I wish to suggest that it is because children possess a system of conceptual structure that functions in terms of logical categories such as proposition, predicate, argument, and quantifier that they are able to impose on expressions like *drink milk* the sort of structure that makes the step Lightfoot discusses trivial. It is not because an expression such as *the milk* is "maximal" in a system of "projections" such as (6) that it is available to combine into a V′ such as *drink the milk* and unavailable to combine into a NP such as **your the milk*, but because its denotation is semantically possible as an argument of the predicate denoted by *drink* and semantically incoherent as the expression defining a domain from which *your* is to pick out a part. Indeed, what makes Determiner + N′ "maximal" is not its internal structure but the fact that its denotation can serve as an argument of a predicate.[1]

Lightfoot says little about how, if at all, logical categories relate to his syntactic category notions. He does not make clear how the traditional category S, the syntactic counterpart of "proposition," is to be interpreted in his allegedly universal system of categories. I conjecture that Lightfoot would reconcile (6) with the one structure (20) in which S appears by allowing INFL as a value of the variable X and interpreting S as INFL′, though some revision of either (6) or (20) would be needed: The Ss of (20) do not consist of what (6b) says an INFL′ could consist of. Does this mean that for Lightfoot a child cannot identify expressions as Ss until he has identified them as projections of something that he has already identified as INFLs? Yes, if category identifications can only be on the basis of phrase structure configurations such as those given in (6), and no, if something else (such as its denoting a proposition) suffices to identify something as a S.

I find the latter a considerably more plausible hypothesis about language, in large part because I take biology more seriously than I think Lightfoot does. Biological structures normally do not arise de novo but evolve out of structures that had different functions.[2] An account of the biology of language increases in plausibility to the extent that the biological structures it posits can be interpreted as having homologs in closely related species and having present or past functions that are not limited to human-style language. I thus rate Lightfoot's purely linguistic conception of linguistic structure as a less plausible hypothesis about the human genotype than the alternative hypothesis that there is a genetically determined scheme of conceptual structure organized in terms of propositions, predicates, arguments, and perhaps some other categories, and that such a scheme of conceptual structure plays a major role in linguistic structure.[3] In McCawley (1982; 1983; 1988), I treat such a system of conceptual/logical categories as determining some of the factors to which syntactic phenomena can be sensitive. (Not all of them: The word-class of the head of an expression – the "X" of (6) – is another such factor, as is a 2-way distinction corresponding roughly to Lightfoot's X/X′ distinction.) My approach allows the syntactically relevant factors to be learned independently of one another (e.g., a child can identify *love* as a 2-place predicate before he identifies it as a V), the approach allows a child to learn rules of syntactic structure in terms of whatever category components he can so far identify (hence prior to the acquisition of the full repertoire of category notions) and it allows him to impose specific syntactic structures on sentence types whose syntax and semantics he has not yet learned (even a child who did not yet understand passive sentences would assimilate *was given money* to the already learned $[_{Y'}YY']$ surface configuration as soon as he had identified *was* as a Y).

The putative category "Specifier" is absent by design from my account of categories. The term "Specifier" was introduced (Chomsky 1970, p. 52) as a cover term for "Determiner" (article, demonstrative, and perhaps some other things), "Aux" (a supposed unit that consisted of the tense marker and whatever auxiliary verbs there were),[4] and degree adjuncts to adjectives; the term has subsequently taken in other things that vary greatly from one work to another. The sole common feature of "Specifiers" mentioned in Chomsky (1970) is that in English they precede phrasal units with which they combine into larger phrasal units. This parallelism in word order is a spurious generalization, since cross-linguistically there is a weak *negative* correlation between (i) whether articles and demonstratives precede the N and (ii) whether auxiliary verbs precede the main verb (McCawley 1982).[5] The arguments offered when proponents of "Specifier" have attempted to justify the category (e.g. Chomsky 1972a, pp. 160–61) have been few in number and scandalously insubstantial; to my knowledge, no serious semantic characterization of "Specifier" has ever been attempted. "Specifier" survives as a device for giving recalcitrant structures a semblance of conformity to the X-bar scheme. It is ludicrous to claim that a category whose very existence is so questionable is genetically determined.

Adherents of the "government and binding" approach to linguistics have no monopoly on the human genotype: All or part of any other conception of linguistic categories and structures can with equal propriety be hypothesized to have a genetic basis. I hope that Lightfoot's target article will stimulate adherents of all approaches to linguistics to worry at least some of the time about how biological structures might be reflected in language acquisition and about how biological evolution might have yielded organisms that are effective acquirers and users of language, but not to form the mistaken impression that GB linguistics has a privileged status in that enterprise.

## NOTES

**1.** This statement refers to the unmarked case, in which predicates take entities and not properties as arguments; there also are predicates with N′ arguments (*Smith has been named chairman of the committee*).

**2.** Jespersen (1922, pp. 412–42), in one of the few accounts of the origin of language that a person knowledgeable in both linguistics and biology can take seriously, develops a fascinating hypothesis about how "preadaptation" not of biological structures but of linguistic material might have figured in the origin of human language.

**3.** I accept Griffin's (1984) conclusion that logical inference is not restricted to the human species. See, however, Anderson and Belnap (1975, pp. 296–300) for a critical discussion of a popular argument that imputes the disjunctive syllogism to the logical repertoire of dogs.

**4.** "Aux" is often confused with "auxiliary verb," even though *must have been* (an "Aux") is no more an auxiliary verb than *lonesome little old* is an adjective. It is widely agreed that no such unit exists, but not so widely agreed that such a consensus exists, since linguists have frequently rejected the concept of "Aux" while retaining the term with a different meaning, namely, "tense or tensed auxiliary verb"; see McCawley (1985) for a discussion of this terminological morass.

**5.** Lightfoot presumably rejects the claimed negative correlation, since he expresses doubts that any languages have N′Spec word order and criticizes Greenberg and Hawkins for "not distinguish[ing] between demonstratives which have the syntax of adjectives (as in Latin) and those which manifest Spec (as in English)." In view of his statement that

Table 1 (Morgan). *Complexity of language input*

| | Degree=0 | Degree=1 | Degree=2+ | N |
|---|---|---|---|---|
| Stage I | 93.28% | 6.38% | 0.33% | 1,504 |
| Stage II | 93.55% | 6.08% | 0.38% | 1,596 |
| Stage III | 89.70% | 9.73% | 0.57% | 1,747 |
| Stage IV | 91.05% | 8.09% | 0.46% | 2,151 |
| Stage V | 87.69% | 11.63% | 0.68% | 1,763 |
| Observed Total | 7,976 | 742 | 43 | 8,761 |
| Projected Total[a] | 3,900,000 | 360,000 | 21,000 | 4,280,000 |

*Note*: Tabulation based on 30 hours of speech transcripts, including two per child per stage. Catenative verbs, such as *wanna* or *gotta*, were not counted as adding a level of embedding.

[a]Projected totals were calculated by extrapolating the hourly rate of speech over eight hours per day, seven days a week, five years.

"in other languages demonstratives and possessives are often adjectives *instead of* specifiers" [emphasis added]), I surmise that is is because Latin demonstratives are inflected somewhat like adjectives that he denies that they "manifest Spec." I see no reason why that should affect the status of *hic*, *ille*, and so on, as "Spec"; in McCawley (1988, p. 194, I argue that part of speech distinctions are orthogonal to the category notions that figure in such terms as "determiner," that is, some "determiners" are adjectives, some are nouns, and some do not belong to any "part of speech." Lightfoot has provided no motivated criteria for the identification of "Spec" that would justify applying it to the prenominal demonstratives of English and Japanese while withholding it from the postnominal demonstratives of Malay and Swahili.

## Learnability considerations and the nature of trigger experiences in language acquisition

James L. Morgan

*Institute of Child Development, University of Minnesota, Minneapolis, MN 55455*

Learnability results utilizing Gold's (1967) framework incorporate assumptions about three components: the amount of information available in language input, the extent of constraints on grammatical hypotheses that the learner may entertain (i.e., the richness of Universal Grammar), and the nature of the psychological mechanisms by which a specific grammar is acquired upon exposure to input. Lightfoot stresses tradeoffs between the first two of these components, but in fact all three are in dynamic balance. Ascertaining the power of one component establishes the combined power of the remaining two.

Of these three components, input is particularly susceptible to direct empirical investigation. The amount of information available is in part a function of syntactic complexity; one metric describing the complexity of input involves tabulating levels of sentence embedding. Applying this metric to speech addressed to Roger Brown's Adam, Eve, and Sarah (Brown 1973; transcripts provided by MacWhinney & Snow 1985), it may be seen that input consists predominantly of simple sentences, as Table 1 shows.

These data suggest an upper limit on the amount of information in input and hence a lower limit on the power of the remaining two components. In particular, any acquisition theory that depends upon the appearance of particular syntactic forms in input – such as those propounded here by Lightfoot, or by Wexler and Culicover (1980) – cannot require input data any more complex than degree-1. This is because in order to ensure sufficient representation of the range of syntactic forms of a given degree of complexity, the number of sentence tokens must be considerably larger than the number of sentence types. However, the number of sentence types expands geometrically as levels of embedding accumulate, so that there are easily thousands, and perhaps hundreds of thousands, of possible degree-2 forms. Even given the patently liberal estimate in Table 1 of the number of degree-2 tokens appearing in input across acquisition (which may well be an order of magnitude too generous), there will be too few such sentences to ensure sufficient representation of degree-2 types.

Power in a theory of acquisition must therefore be sought somewhere other than in the syntactic complexity of input. One may seek additional dimensions by which information may be conveyed in input (Morgan 1986), or alternative learning strategies (Osherson et al. 1986), or more powerful constraints on Universal Grammar (Lightfoot). In specifying a rich theory of grammar as the basis for claims of learnability from very simple input, however, Lightfoot fails to articulate clearly the third component of his model. With respect to the specification of the pyschological mechanisms of acquisition, two questions are of particular importance. The first concerns how children come to represent input in an appropriate fashion; the second concerns the nature of the relationship between triggering experiences and the parameters whose settings are triggered.

Several of the arguments adduced by Lightfoot in support of his degree-0 hypothesis depend crucially on particular representations of input. Consider, for example, his account of differences in government between English and Dutch (see section 5.2). English does not allow the subject of an embedded sentence to be *wh*-questioned if an overt complementizer is present; Dutch does. The explanation advanced for this is that the presence of the complementizer blocks government in English but not in Dutch. Lightfoot suggests that sentences like (1) (= his (30)) provide degree-0 evidence that, in Dutch, government by a *wh* pronoun is not blocked by the presence of additional lexical material in Comp. This argument thus depends upon the auxiliary *heeft* being moved into Comp. However, Lightfoot acknowledges the possibility that the auxiliary may not move to this position (see note 12); (2) is a possible alternative sentence structure. If children represent *wie heeft het boek gelezen* as (2) rather than as (1), Lightfoot's degree-0 argument fails to go through in this case.

1. $_{Comp}$[wie$_i$ heeft$_j$], $_S$[e$_i$ het boek gelezen e$_j$]
   who has the book read
   "Who has read the book?"
2. $_{Comp}$[wie$_i$] $_S$[heeft$_j$ e$_i$ het boek gelezen e$_j$]

Triggers, as portrayed by Lightfoot, are bracketed and labeled sentence structures. Children, however, hear strings of sounds; these structures must somehow be mentally constructed by the child. Presumably, Universal Grammar will constrain the possibilities for the representation of any input datum. Even so, alternative representations will often be possible, as illustrated. Theoretical justifications for particular representations, independent of degree-0 learnability considerations, should be provided, as should explanations of how children must arrive at these representations.

The sort of trigger-parameter relationship that Lightfoot intends is not clear. In every example, the triggering data provide evidence contradicting the current setting of a parameter. These examples are thus consistent with an error-detection learning model of the sort offered by Wexler and Culicover. Such a model enforces associations between the content of input and the nature of modifications to the learner's internal grammar. But Lightfoot apparently has something else in mind: "The error detection procedure reflects a more fundamental difference in orientation. . . . I have adopted the recent parameter-setting model, . . . [which] specifies certain option points which are set on exposure to particular data-sets" (note 8).

To illustrate the sort of problem that may arise upon forsaking an error-detection model, consider a recent proposal by Hyams (1987). Hyams's concern is the Null Subject parameter, which determines whether sentence subjects are obligatory (as in English) or optional (as in Italian or Chinese). Languages allowing optional subjects have uniform inflectional marking of verbs – Italian has inflections for all combinations of person, number, and tense; Chinese has no inflections whatsoever. In contrast, languages with obligatory subjects have nonuniform inflectional marking. Hyams proposes that optionality is the unmarked setting of the parameter, with the marked setting triggered by morphological nonuniformity. Thus the receipt of any present tense paradigm (e.g., *I run, you run, he runs*) is sufficient to indicate that English has obligatory subjects, despite the fact that these data are consistent with the unmarked setting of the parameter. Hyams may have tapped some deep systemic relationship in language, but on the face of things the connection between the Null Subject parameter and its putative trigger is quite arbitrary.

If arbitrary trigger-parameter relationships are admitted, it will always be possible to find simple triggers (these need not be consistent across languages, or even across learners), and claims that triggers must comprise input of a given degree of simplicity will be unfalsifiable. Unfortunately, alternative sources of constraints on these relationships appear to be unavailable. Theoretically, one could draw analogies in learning across cognitive domains, except that the properties of Universal Grammar are commonly taken to be specific to the "language faculty" (cf. Chomsky 1986; 1988). Empirically, one could observe relations between input and acquisition in development, except that failures to find expected trigger-parameter relations may always be attributed to maturational factors.

In sum, Lightfoot is right in insisting that acquisition must proceed on the basis of robust data; clearly this comprises simple sentences. Lightfoot is also right in arguing that acquisition theories should account for language variation. Comprehensive evaluation of his acquisition model, however, must await more explicit specification of the psychological processes involved.

## On *one* as an anaphor

Stephen Neale

*Department of Philosophy, Princeton University, Princeton, NJ 08544*
**Electronic mail:** *neale@confidence.princeton.edu*

Lightfoot takes it as uncontroversial that "any noun phrase [NP] that occurs in English, and thus any noun phrase that an English-speaking child will hear" can be generated by either of two incompatible sets of phrase structure rules. For example, the NP *an old man from the city* might be analyzed as either (i) a Specifier (Determiner) *an* concatenated with *old man from the city* (an N′, a phrasal category intermediate between a common noun and a full NP), or (ii) an NP *an old man* concatenated with a prepositional phrase (PP) *from the city*. Two pieces of evidence in favor of (i) emerge from investigation of the behavior of *one* when used as some sort of anaphor. First, there is the contrast between Lightfoot's examples (4.a) and (5.a):

(4) a. *The student of physics is older than the one of chemistry
b. The student from New York is older than the one from Los Angeles

where *one* is supposed to be anaphoric on *a student*. Second, there is the fact that (5.a) is ambiguous according as *one* is understood as anaphoric on *suit* or on *old suit*.

(5) a. He wants an old suit but he already has the only one I own

At first sight the contrast in (4) is puzzling, but Lightfoot provides a plausible account of it. In (4.a) the common noun *student* has as its sister node the PP *of physics*, and both are immediately dominated by an N′. In (4.b), by contrast, the common noun *student* is immediately dominated by the N′ *student* which is the sister of the PP. Lightfoot's hypothesis is that *one* can "refer back to an N′" but not to a common noun. This licenses (4.b) but not (4.a), and permits *one* in (5.a) to "refer back to" either *suit* or *old suit*, both of which are N′s.

The important conclusion we are to draw from this is that children must attain a grammar that analyzes NPs in the manner of (i) rather than (ii) on a "nonexperiential basis." The child's primary linguistic data are simply too impoverished to determine the particular grammar the child obtains; they fail to provide the information necessary to induce certain principles and generalizations characteristic of that grammar. The positive proposal is that Universal Grammar – the linguistic genotype – contains the information that (a) an NP consists of a Specifier and an N′, and (b) an N′ consists of a head (N′ or N) and a complement. (The linear order of occurrence is presumed to be determined on the basis of experience.) The reason given for this claim is that although language learners might encounter any NP that could occur in English, they are not "systematically informed" that strings like (4.a) are not normally used and that (5.a) is ambiguous.

I do not wish to challenge the substance of this interesting claim. However, it seems to me that a certain amount of semantical clarification is called for in Lightfoot's discussion of the anaphor *one* before the syntactic claim underlying the psychological claim can be interpreted. In the above examples it would be perfectly fine to say that *one* is *anaphoric on* an N′. But what Lightfoot says is that *one* "refers to" or has as its "referent" a "preceding N′" (note 1), and this is highly problematic. First, since reference is a word/world relation, if *one* referred it would refer to some nonlinguistic object and not the expression upon which it is anaphoric. What Lightfoot should say is simply that *one* is anaphoric on an N′. Indeed, talk of *reference* is quite out of place here. It is certainly not true that *one* has to be coreferential with its antecedent or anything of that sort. (*Bill bought a car yesterday and Mary bought one today* can be true even if Mary does not buy the car Bill bought.) Rather, *one* is best viewed as what Peter Geach (1962) has called a "pronoun of laziness," that is, a pronoun which, from a semantical perspective at least, stands in lieu of a repeated occurrence of its antecedent. (Notice also that *one* can precede its "antecedent" as in *seeing Doris buy one yesterday encouraged Mary to buy a wig today*.)

Having made this move a second problem emerges. If *one* is a pronoun of laziness, then on Lightfoot's account it should stand

in for an N′. But, as we just saw, *one* may stand in for a full NP like *a car*. (See also Lightfoot's example 15.) Lightfoot could now suggest that although *one* may, on occasion, *stand in* for a full NP, it is actually an N′ within that NP upon which it is *anaphoric*; but the resulting mismatch between the syntax and semantics of the type of anaphora under consideration would appear to be quite unmotivated. At the very least, then, we should accept that *one* may be anaphoric on N's and NPS. (Notice that this does not have to drag reference back into the picture since the most plausible account of indefinite descriptions treats them as quantified noun phrases rather than as referring expressions; see Russell 1919, ch. 16; Ludlow & Neale 1988.)

Finally, it should be pointed out that Lightfoot is overstating the case when he claims that "*any* [my emphasis] noun phrase that occurs in English" can be generated in the manner of either (i) or (ii) above. Proper names and pronouns (including the demonstrative pronouns *this* and *that*) are most certainly noun phrases, but since they have no internal structure they cannot be generated in either manner. To claim that names and pronouns may be N's dominated by NPs would of course undermine the view that *one* may be anaphoric on an N′ since *one* can never be anaphoric on a name or pronoun.

## Two perspectives on learnability

William O'Grady

*Department of Linguistics, University of Calgary, Calgary, Canada T2N 1N4*

David Lightfoot's target article takes as its starting point the assumption that linguistic development depends on the interaction of inborn cognitive structures with selected types of experience, a view that constitutes the consensus in cognitive science today. Indeed, as Osherson and Wasow (1976, p. 208) note, the only real question facing linguists is "whether innate structure has significant components that subserve the development of no other faculty than language." For the sake of convenience, let us refer to the view that there are such components as "special nativism," reserving the term "general nativism" for the view that the inborn cognitive structures relevant to linguistic development are not specific to the language faculty.

Lightfoot, of course, is a strong proponent of special nativism. Not only does he believe that there are inborn cognitive structures, he believes that they include faculty-specific components such as (1) and (2).

(1) XP → Spec., X′ (example 6)
X′ → $\begin{Bmatrix} X' \\ X \end{Bmatrix}$, (YP)

(2) *One* refers to an N′. (Sect. 4, para. 8 & note 1; Lightfoot 1982, p. 56)

My purpose in this brief commentary is to challenge this type of nativism. Because of space limitations I will restrict my remarks to just one of the phenomena discussed in the target article – the contrast between *the one from Los Angeles* as in Lightfoot's (4b) and **the one of chemistry* as in his (4a). My goal will be to show that this contrast does not demonstrate the need for an N′ category or for the inborn mechanisms in (1) and (2) above.

Like Lightfoot, I assume that an inborn acquisition mechanism (AM) is needed to construct the grammar of a human language. Unlike Lightfoot, however, I envision an AM that does not contain inborn principles specific to the language faculty. To illustrate this approach, I will outline one component of this alternative AM and show how it can help explain the distribution of *one*.

In what follows I will assume that the AM contains a "propositional module" that provides a means to represent propositions in an innate "language of thought."[1] Among other things, this module provides a representation of word meanings and of any ontological or "thematic" roles that they entail. The lexical semantics of *destroy*, for example, entails an entity that commits an act of destruction (an agent) and an entity that is destroyed (a theme), *from* entails a source, and *student* an optional theme (a thing studied).

(3) *destroy:* ⟨agent, theme⟩
*from:* ⟨source⟩
*student:* (⟨theme⟩)

I also assume that the propositional module includes information about the association of these semantically determined roles with particular arguments. Hence:

(4) a. Harry destroyed the book
[agent] [theme]
b. student from NY
[source]

Finally, I will assume that an inborn semantic coherence requirement demands that each referring expression bear a thematic role (e.g., Kaplan & Bresnan 1982, p. 212). This requirement is met in sentences such as *Harry destroyed the book*, in which each NP is associated with a role (agent and theme, respectively), but not in **Harry arrived the book*, in which there is no role for the second NP.

Returning now to the problem of *one*, there is no violation of the coherence requirement in phrases of the type *student of physics* or *student from NY*. In the former case, *physics* is associated with the theme role determined by the lexical semantics of *student* (see 3)), whereas *NY* is associated with the source role entailed by *from*.

(5) a. student of physics
[theme]
b. student from NY
[source]

Why, then, is *one* substitution permitted only in the second case?

(6) a. *the one of physics
b. the one from NY

This contrast follows from a simple and easily observable fact about *one*, namely:

(7) *One* has no lexical content.

Since *one* has no lexical content, it cannot determine a thematic role. In order to satisfy the coherence requirement, the NPs in (6) must therefore receive a thematic role from something other than *one*. This happens in *the one from NY*, where the preposition has enough lexical content to assign a source role. In the ungrammatical **the one of physics*, in contrast, the preposition is semantically empty and therefore incapable of assigning a thematic role.[2] Given that *one* also does not specify a thematic role, *physics* runs afoul of the coherence requirement and the phrase is therefore ill-formed.

From the point of view of learnability, this analysis offers some interesting contrasts with Lightfoot's proposal. Whereas Lightfoot must stipulate that there is an N′ category and that *one* refers back to this category type, I have only to stipulate that *one* lacks lexical content. Moreover, whereas Lightfoot's stipulations are not learnable, children should be able to determine that *one*, unlike *student*, lacks lexical content. In fact, even the strongest proponents of special nativism recognize that children must be able to learn the idiosyncratic properties of the lexical items in their language.

I have been compelled to focus here on a small issue, ignoring many of the other points raised by Lightfoot (including some that I agree with, such as the desirability of making the AM

sensitive to very simple forms of experience). Furthermore, although I have tried to provide an analysis consistent with general nativism, space has not permitted discussion of many details, including the (N′-free) syntactic structure that should be assigned to "*one* phrases." I have likewise been unable to consider any of the other mechanisms Lightfoot proposes (e.g. constraints on multiple *wh* structures). Of course, for many phenomena, there are simply no proposals to make at this time. The principles proposed within contemporary special nativist theories (e.g. Lightfoot's (27)) were not discovered at an early point in the development of that research program, and there is no reason to expect their counterparts within general nativism to be uncovered quickly either.

Given the long established pattern within linguistics of looking first to abstract syntactic notions for the description of both new and traditional problems, it is clear that analyses challenging special nativism will not simply spring forth. If they are to be discovered, they will have to be very deliberately pursued. Moreover, if the early work on special nativism (as exemplified by transformational grammar) is in any way typical of how theories emerge in the cognitive sciences, we can expect preliminary proposals to be often incomplete and to lead to many disappointing dead ends. Counterbalancing these inevitable difficulties, however, is the possibility of a significant advance in our understanding of the human language faculty – a possibility that may well make this line of inquiry worth pursuing.

NOTES

1. I omit discussion of the computational module, which is concerned with the formation of structural representations, and of the hypothesis formation module, which uses cross-faculty "learning" mechanisms such as generalization (constrained by a conservatism law) and modus ponens to acquire particular principles. For discussion of these issues, see O'Grady (1987 and forthcoming).

2. Where *of* retains a modicum of literal meaning and designates a "content" relation as in *pitcher of milk*, it can assign a thematic role (hence the acceptability of . . . *and one of water*).

## On the format for parameters

Luigi Rizzi

*Faculté des Lettres, Université de Genève, CH-1211 Geneva 4, Switzerland*

**Electronic mail:** *rizzi@cgeugesi.bitnet*

The target article has the merit of making explicit an assumption that is rarely spelled out but absolutely crucial for the parametric approach if the latter is to be construed as part of an abstract model of language acquisition: Values of parameters must be easy to select on the basis of simple evidence readily available to the language learner. If this obvious feasibility constraint is not met, a parametric analysis will not adequately address the acquisition problem, no matter how structured our assumptions about the predetermined linguistic mechanisms may be.

If there is no doubt that checking the feasibility of parameter setting is of fundamental importance, it is less obvious that this goal requires a principled approach. The need for a special principle limiting the search domain for parameter fixing is not self-evident and its conceptual status raises questions. What could be the theoretical nature of the constraint that the search space is limited to "a clause plus a little bit"? Surely such a principle could not be part of Universal Grammar (UG) in the same sense in which familiar principles (Subjacency, ECP, etc.) are. Moreover, it is not obvious that a "degree-0" constraint could be construed as a principle of learning theory, entering into the acquisition process in interaction with UG, but not belonging to UG proper. Compare this constraint with a reasonable candidate for a principle of learning theory: the Subset principle (Berwick 1985). Whereas the latter is abstract enough to be plausibly construed as a component of a general learning theory, applying across cognitive systems (i.e., whenever a learning system is confronted with an option that would increase the class of objects that the system to be attained can deal with, the option is taken only if experience requires it), the degree-0 constraint, as stated, looks too concrete and modality specific to be ascribed to a general learning theory. In sum, the conceptual status of the constraint is not transparent.

I would like to explore briefly a possible way to deal with the conceptual problem and, at the same time, to put an even more severe restriction on the search space for parameter fixing, fully in the spirit of Lightfoot's proposal. The idea that a very limited amount of structure suffices for parameter setting can be essentially right but need not be stated as a separate principle: It can simply fall out as a consequence of the format that UG offers for parameters. Most well-established cases of parameters have to do with properties of heads: Either they are associated with specific lexical items (as Hagit Borer pointed out), or with heads of certain kinds, or with heads in general (i.e., the ordering between heads and complements, the licensing of null pronominals by Infl or other heads [Rizzi 1986], various parametrized properties of selection and case assignment, etc.). If the space available for parametrization is indeed restricted to properties of heads, it is natural to expect that the evidence needed for the fixation of each parameter will not exceed the normal sphere of influence of a head: its government domain. The fixation of parameter P will (at most) require inspecting the amount of structure governed by the head that the parameter is related to.

It should be noted that this way of looking at the problem puts a further restriction on the search space for parameter fixing with respect to Lightfoot's formulation: neither sentential subordination, nor also phrasal subordination (hence, subordination in general) need be accessed. This higher restrictiveness seems to be justified: For instance, there is no plausible case of a parameter requiring more than one level of nominal embedding to be fixed. And the "little bit" of subordination that is required for parameter fixing can now be characterized in a nonarbitrary way: It involves cases in which a head extends its government domain to material belonging to the subordinate structure, as in the *believe* case and in the cases involving a subordinate Comp position mentioned by Lightfoot.

One might object that our proposal is too restrictive. In particular, the S/S′ case falls out of the picture, while it can be integrated within an approach expressed in terms of sentential subordination, as Lightfoot shows. Perhaps it is not undesirable to set this case aside, however. The S/S′ case is of some historical significance, as it was the first concrete instance of language variation treated in terms of the parametric approach; as such, it gave impulse to subsequent research in this area. But, seen in retrospect, this case looks quite atypical – certainly not representative of what is now known about parameters. First, no other major property seems to be clearly related to the extractability from wh-islands (the facts of extraction from NP in Romance are quite complex, and in any event could not be related to the wh-island facts in a principled approach to subjacency such as Chomsky 1986, so one does not find the clustering of properties occurring together that makes the parametric approach interesting. Second, judgments vary considerably among speakers (e.g., Grimshaw 1986, describes an English dialect which essentially corresponds to the description of Italian in Rizzi 1982, ch. 2) and there seems to be no neat distinction between grammatical systems, as with other parameters. Third, unlike the major familiar cases of parameters, this one does not seem to be reducible to a property of a head. (In Chomsky 1986, there is a recent assessment of the issue.)

I would like to conclude this commentary by pointing out that Lightfoot's approach, whatever the theoretical status of the proposed constraint will ultimately turn out to be, has an important heuristic value. First of all, it has the effect of excluding a large class of unattested, bizarre parameters. Even

more important, it enforces a reanalysis of conceivable parameters along directions that tighten the deductive structure of the system. A case in point is provided by the existence of long distance anaphors (Manzini & Wexler 1986). Whereas anaphors such as *himself* in English are clause-bound, some languages allow anaphors to have their antecedents in a higher clause (e.g. Icelandic, Anderson 1986):

(1) Jón segir ad María elski sig
John says that Mary loves+SUBJ himself

One could think that, in order to determine whether his language has long distance anaphors, the language learner should hear such sentences as (1), in which the anaphor lacks an antecedent in its simple clause. But this would be in violation of degree-0 learnability (and of the refinement we have suggested). So, if we hold the latter constant, we must look further to explain the acquisition of long distance anaphors: The fact that *sig* is long distance in Icelandic and *himself* is not in English must be related to some independently observable property of the two elements or of the two grammatical systems. In fact, there is a plausible candidate for the case at issue. Pica (1987) has noticed that (nonclitic) morphologically simple reflexives like *sig* tend to be long distance, whereas morphologically complex anaphors like *himself* in general are clause bound, a generalization that he attempts to explain in terms of a theory of anaphor movement in Logical Form. If this approach is on the right track no conflict with degree-0 learnability arises. A simple inspection of the form of the anaphor provides sufficient evidence to determine its local or long distance status.

Even if many details remain to be settled (for instance, the fact that in different languages long distance anaphors seem to be sensitive to different kinds of tense/mood distinctions in the clauses intervening between the antecedent and the anaphor), the positive heuristic value of degree-0 learnability is clearly illustrated by this example. In enforcing the reanalysis of cases of this kind, Lightfoot's approach makes it harder to stop too early in the analysis of cross-linguistic variation and favors seeking out new correlations and deepening the deductive connections in the theory of grammar.

## Language acquisition: Dubious assumptions and a specious explanatory principle

I. M. Schlesinger
*Department of Psychology, Hebrew University, Jerusalem 91905, Israel*

Lightfoot's interesting target article is based on theoretical premises which he shares with writers on language acquisition working in the generative grammar tradition. My comments are directed at the weaknesses of these theoretical presuppositions, not at Lightfoot's ingenious analyses based on them.

The case for what Lightfoot calls the "selective" model of human language capacity rests on the claim that the child eventually acquires knowledge that is "not part of the input," that is, could not have been derived from linguistic experience. However, in arguing for this claim generativists make two implicit assumptions, both unwarranted.

One assumption is that the child is informed only about the correctness or (according to one view, not accepted by Lightfoot) incorrectness of sentences. No account is taken of the possibility that the child acquires information about the frequency of a certain type of construction. (On the human propensity to register such information see, e.g., Hasher & Zacks 1984.) Frequency information goes a long way toward accounting for language acquisition. Braine (1988) describes a learning model that is sensitive to such frequency information and disposes of the problems of "degenerate input" and of the unavailability of negative feedback. Although the basic idea of this model was outlined long ago (Braine 1971), writers on generative grammar have never, to my knowledge, discussed this alternative.

Generativists claim that unless one posits innate knowledge there is no answer to the question of how the child arrives at abstract grammatical constructs (such as N′ in Lightfoot's example). This argument is based on another implicit assumption, namely, that the analysis made by the child is at the abstract level found appropriate for linguistic analysis (e.g., Lightfoot's (2.a)); that is, that the child, like the linguist, attains the rule that explains the maximum data. This has gone unquestioned ever since Chomsky (1965, p. 30 ff) postulated an evaluation measure by which the child selects the simplest, most economical description of the data.

But why should the child not entertain hypotheses at a less general level, at least in the early stages of learning? There is no reason for the child to aim immediately at abstractions at the highest level when the data are amenable to a simpler, more "local," treatment; see Schlesinger (1982, pp. 31–34) for an example. Eventually the child may then subsume the lower-level rules under more comprehensive ones. Or he may not; for the organization of the data most efficient in acquisition and for the processing mechanism may differ from the most economical linguistic description. The second implicit assumption, then, is also entirely unwarranted.

Let us now, for the sake of the argument, take these assumptions for granted and turn to the generativist solution to the "poverty-of-the-stimulus" problem: the postulation of "genotypic" (i.e., innate) information. This move might have some merit if it gave a general solution to the problem of knowledge that cannot be derived from the input – according to the above assumptions. Actually, generativists discuss only examples suggested by the government/binding model, and these represent only a very small subset of the knowledge we eventually acquire. Innate constructs are of no avail in many other cases, which are explained easily and quite naturally in terms of general cognitive factors. This makes the generativist solution extremely unparsimonious.

For instance, that the asterisked sentences in (1)–(6) are unacceptable cannot be derived from the linguistic input any more than the cases discussed by Lightfoot. But no proposal has been forthcoming as to how the relevant knowledge might be attained with the help of "genotypic" information (although examples like these have been in the published literature for much over a decade).

(1) a. The car scraped the tree with the front wheel rim.
b. *The wind broke the window pane with the twisted branch.
(2) a. The crane picked up the box.
b. *The fork picked up the potato.
(3) a. The janitor's key opened the door for him.
b. *The janitor's key opened the door for him without delay.
(4) a. The tent puts up in less than five minutes.
b. *The tent puts up in the backyard.
(5) a. This shirt washes easily, because it is made of synthetic fabric.
b. *This shirt washes easily, because I have a lot of time.
(6) a. As a single woman, Jill used to bolt her front door, and it was lucky for her that she did so/that.
b. *As a baby, Jill resembled her older sister, and it was lucky for her that she did so/that.
c. *As a bachelor, John used to admire the Swingle Singers, and it is strange that he did so/that.

An explanation of (1)–(3) is proposed in Schlesinger (1985; 1988) in terms of semantic factors (subjects are conceived of as agents, which makes the above (b)-sentences unacceptable). Examples (4)–(5) are adapted from Van Oosten (1977), who offers a similar explanation (the subject position implies respon-

sibility). The semantics of the verb category (dynamic vs. stative) is involved in (6) (for details see Quirk et al. 1972). The b-sentences in (1)–(6) can be generated from the same rules as the corresponding a-sentences. Children cannot learn from the input that the former are unacceptable, and they do not need to: They avoid applying the general rule because of the unintended connotations of the (b)-sentences. In other cases, relative processing load (which has been shown to determine partly the relative acceptability of Dative Movement; see Schlesinger 1977) may account for unacceptability.

The generativist approach, then, merely provides us with a special-purpose solution for certain specific problems; it is inapplicable on a wider scale. Since pragmatic and semantic factors and considerations of processing load do not involve postulating additional entities, it would be good strategy to explore these factors and see how far they can take us before deciding to populate the human "genotype" densely with putatively universal constructs. Unless such a course is adopted, our language acquisition theory will be too homuncular.

ACKNOWLEDGMENT

I thank Yonata Levy for helpful comments on a draft of these comments.

## Data on language input: Incomprehensible omission indeed!

Catherine E. Snow[a] and Michael Tomasello[b]

*[a]Graduate School of Education, Harvard University, Cambridge, MA 02138 and [b]Department of Psychology, Emory University, Atlanta, GA 30322*

**Electronic mail:** *[a]Snowcat@hugse1.bitnet and [b]psymt@emory.u1.bitnet*

Lightfoot's stated goal is to focus attention on the role of primary linguistic data (PLD) in the child's acquisition of language. He argues that the total neglect of this problem in the learnability literature constitutes an "incomprehensible omission." He then provides an analysis of the role of PLD that neglects virtually all of the empirical research on PLD – an even more incomprehensible omission from our point of view. As researchers who have contributed to this research, we thus find ourselves endorsing Lightfoot's goal, but puzzled about the limited means he uses in its pursuit. In this commentary, we will make some suggestions about additional sources of information that might be used in addressing the very important issues Lightfoot raises. We will argue that one reason these sources are neglected in this and other works in the learnability paradigm is the reliance on the metaphors of "triggering" and "parameter setting."

The basic question Lightfoot formulates is: What differentiates the parts of the PLD that constitute a basis for children's language learning from those that do not? He very correctly points out that not all the experiences the child has, nor all the sentences a child hears, are used as a basis for generalizations about language. This is a point that has been made again and again in the more psychologically based approaches to language acquisition (e.g., Nelson 1982; Snow 1972; 1977). The correctness of this point is assumed by virtually everyone doing empirical research on PLD (as well as by most other developmental psychologists) and is in fact one reason researchers in this area avoid the term "input" in favor of terms such as "intake" and "uptake." Thus, for example, Lightfoot's imputation that Wells (1981) was so naive as to think that everything uttered in a child's hearing constituted usable PLD is quite wrong. The problem remains, though, to offer a principled basis for distinguishing between usable and nonusable PLD.

Lightfoot provides one promising approach to the problem. In a very interesting analysis, he demonstrates how specific cases of historical language change provide a basis for hypotheses about the selective reanalysis of certain types of utterances, and thus strong evidence that those utterances and novel analyses of them constituted usable PLD for children at some particular time period. While recognizing the value of such historical evidence, we must also point out that many additional sources of evidence about the nature of usable PLD are available here and now.

Although Lightfoot does not say so explicitly, it is clear that before an adult sentence can act as PLD for the parameter-setting mechanism, the child must first provide that sentence with an appropriate syntactic analysis. A parameter cannot be set by a string of English or Japanese words; it must be set by some language-independent structural description of such a string. Thus, for a sentence such as "His belief that Susan left guided his actions" to act as appropriate PLD the child must, among other things, classify "belief" as a noun which is part of the subject NP (and not, for example, as a verb) and must designate "guided" (and not "left") as the main verb of the sentence. Though again Lightfoot is not explicit on this point, many learnability theorists (e.g., Pinker 1984) argue that in many instances the child's syntactic analyses can only be made with the assistance of semantic information. It would seem then, for example, that the only way for children to provide appropriate structural descriptions of passive sentences is either to restrict their attempts to pragmatically irreversible sentences or to hear reversible passives such as "The cow was kicked by the horse" in contexts which make the intended meaning clear.

The first step in distinguishing usable from nonusable PLD is to determine which aspects of adult speech children are able to process, both semantically and syntactically, at various developmental periods. There is a large psychological literature on precisely this point. A variety of studies have established the specific cognitive and linguistic skills prerequisite to children's comprehension of particular linguistic structures, from English word-order conventions at two years of age (Huttenlocher 1974; Sachs & Truswell 1978) to passive sentences and embedded clauses at ages five to seven (Horgan 1978; Maratsos 1974, Slobin 1968). If the question is usability, many of the answers are in this literature.

Another extremely rich source of data on which aspects of PLD children are processing may be found in the communicative-interactive analyses normally eschewed by learnability theorists. In such analyses, children on many occasions confirm directly that they have processed a piece of adult language and also provide evidence that some discourse conditions make such processing easier. First, children sometimes respond to a novel structure in adult speech in a way that could only indicate comprehension (with other sources of sentence comprehension, e.g. nonlinguistic cues, unavailable). This seems to be made easier by adult language that is on the topic of the child's immediately preceding utterance (Snow et al. 1987; Snow, in press). Second, children often imitate novel and progressive adult forms (and this is true despite the single, worn-out putative counterexample from McNeill [1966], which only makes the point that children can't imitate everything; for examples of children using imitations to expand their own systems, see Clark 1977; 1978; Snow 1981; 1983). Most important, children are especially likely to imitate when the adult has just recast their immature linguistic form into a mature adult form (Bohannon & Stanowicz 1988). Third, children often spontaneously produce at some later time a novel structure addressed to them by an adult on an earlier occasion. Experimental studies controlling PLD and measuring the resultant learning have established, for example, that the introduction of new forms in certain discourse contexts, especially recasts, facilitates the child's acquisition of these forms (Nelson 1982).

The important point from our perspective is that the usability of a particular piece of PLD – and thus its availability to act as a trigger – relies on the child's comprehension and analysis of it. This in turn depends on (i) the developmental level of the child, both cognitively and linguistically, and (ii) the communicative-interactive conditions surrounding adult use of that structure. The child language literature is replete with empirical research

in both of these areas. We find the omission of references to this highly relevant literature in the target article incomprehensible; worse, it undermines serious attempts to consider how children actually use the PLD as a source of information about the linguistic system.

A primary reason for Lightfoot's neglect of empirical research on child language acquisition may be his reliance on the "triggering" and "parameter-setting" metaphors (deriving ultimately from an incorrect view of the biological bases of language, we would argue, but that is another story). The term "trigger" suggests that a single event causes the relevant change in the child's linguistic system irreversibly and instantaneously. This is not universally wrong, and in fact a similar notion is implied by Keith Nelson's (1987) "rare-event model," which likewise proposes that a single interactive event at the right moment can induce significant change in the child's language system. The difference between "triggering" and "a rare event" is that Nelson's model includes considerable attention to the developmental achievements that are prerequisite to the rare-event effects, some of which may result from a long, slow, laborious process of incremental change that is a precondition to the crucial utterance's targeting the right parameter.

We must also emphasize that insofar as the triggering metaphor implies error-free performance after the parameter is set, it is just plain wrong for many if not most of the structures of early child language (nor is the invocation of a competence/performance distinction an adequate response to the ubiquitously meandering character of developmental processes). In many domains of language acquisition, development is more correctly characterized as gradual movement toward adult-like performance, often followed by a period of increased errors that signal reanalysis and/or reorganization, succeeded again by more adult-like performance (e.g., Bowerman 1982). The data from children learning language simply do not support the view that learning a linguistic structure is as simple and irreversible as hitting a switch.

Another implication of the triggering metaphor that seems particularly questionable to those of us who have studied transcripts of children talking is the assumption that restructurings in the child's system are solely the result of comprehension events. Transcripts reveal the enormous amounts of effort children put into producing some utterances, particularly those at the edge of their own linguistic systems. It seems clear that in many cases children are "learning by doing," that is, the structural analysis that provides the basis for the so-called triggering experience may come as children struggle to express themselves in adult-like ways. Testing this possibility requires supplementing the research strategy of making assumptions about sentences children must have heard by studying transcripts of what they actually say.

Perhaps the most disturbing aspect of the triggering and parameter-setting metaphors for the developmental psychologist who has taken the trouble to document the process of language acquisition is the degree to which they imply that the child is a passive organism – a target or set of targets to be hit by the utterances produced by adults. These metaphors fail entirely to capture the most salient feature of the language-learning child: active industriousness. Children work very hard in the process of comprehending and producing utterances. A close look at individual children acquiring particular structures (e.g., the use of prepositions or the past tense of regular verbs) often reveals days, weeks, or even months during which they hear and struggle to comprehend dozens or even hundreds of relevant adult examples, and during which they struggle on numerous occasions, interspersing both successes and failures, to produce the correct form themselves (see Tomasello 1987; forthcoming). All of these variations and vacillations are merely noise to the parameter switchboard; that may be why learnability theorists neglect them.

Ever since Chomsky severed competence from performance and restricted the definition of language to syntax alone, there has been a profound lack of communication between linguistically and psychologically oriented developmental psycholinguists; for the most part we are not even talking about the same thing. By focusing on PLD from the learnability perspective Lightfoot has potentially narrowed the rift. He demonstrates quite clearly that information about PLD is crucial for researchers of both persuasions because our characterizations of PLD and the language acquisition mechanism are interdependent: Every hypothesis about a facilitative aspect of the child's social interaction is also a hypothesis about the child's mind. We believe that the concept of PLD used by Lightfoot and other learnability theorists is too narrow, that it neglects the information children gain from comparing their own utterances to those of adults in the same discourse context, and that it ignores important information about how developmental level interacts with the use and acquisition of language. But at least we are beginning to talk about the same thing.

## What's a trigger?

Edward P. Stabler, Jr.

*Department of Computer Science, University of Western Ontario, London, Canada N6A 5B7*

**Electronic mail:** *stabler@uwovax.bitnet*

The development of the principles and parameters approach in GB (government binding) theory makes the learning problem look more tractable than it did in the framework of the Standard Theory. Rather than learning a set of transformations with restrictions on the order of their application, the language learner need only set a finite number of parameters. Of course, a solution to the language acquisition problem must account for the fact that a child can learn any human language, so some parameter setting must define (core properties of) each human language. And the process should be "feasible" in the sense that the determination of the appropriate setting must be achievable "within the given constraints of time and access, and with the range of observed uniformity of output" (Chomsky 1965, p. 54). Within this framework, Lightfoot proposes that relatively simple data suffice to determine parameter settings. Specifically, no data containing "embedded Domains" is required: "Given certain formulations of UG, the relevant parameters can be fixed on the basis of data in nonembedded Domains." The confirmation of such a proposal is a step toward confirming the feasibility of the principles and parameters approach, since it shows that no implausible assumptions need be made about the complexity of the structures a child needs.

Lightfoot's claims are appealing, but the arguments for them rely on an oversimplified account of parameter setting. I think that Lightfoot will agree that at least a few promissory notes are involved. His account is subject to the following familiar problems:

1. The analyses used to support the account are controversial and unstable.

2. If parameter settings interact, the account of parameter setting must be elaborated to ensure that the account does not run afoul of the subset principle.

I would like, however, to focus on two other interrelated problems that are, I think, more serious. These do not undermine the familiar idea that it could be valuable to use learnability considerations to guide the development of linguistic theory, but they do raise doubts about the claimed sufficiency of nonembedded or "degree-0" Domains:

3. Lightfoot is not clear about the kind of data the learner is presumed to have access to. The examples of triggers suggest an implausible view; nor is it clear that more plausible assumptions will still allow a defense of the degree-0 hypothesis.

4. The assumptions about parameter setting that are suggested by Lightfoot are empirically untenable. Triggers must have some properties which distinguish them from other sorts of linguistic data, but these properties are neither specified nor shown to be possessed by the degree-0 examples.

I consider the latter two points in turn.

***1. What is in the primary linguistic data?*** Each parameter of UG is presumed to have a finite number of settings, ordered by markedness. Let the symbol $\vec{p}_i$ represent an assignment $i$ of values to all of the parameters. Then UG($\vec{p}_i$) represents a grammar. The simplest idea is perhaps that the language learner uses his current parameter settings UG($\vec{p}_i$) until prompted to adopt more marked settings UG($\vec{p}_j$) by part of the primary linguistic data (PLD). Given the PLD and a current grammar UG($\vec{p}_i$), a change of parameter settings might be prompted in something like the following situation:

(i) the primary linguistic data (PLD) contain some structure S that is not well-formed according to UG($\vec{p}_i$),

(ii) S is well-formed when some of the current parameters are set to more marked values, yielding a different grammar UG($\vec{p}_j$).

I will call this the "standard account of parameter setting." In the situation just described, PLD, and especially S, is the "trigger" for the development of the grammar from UG($\vec{p}_i$) to UG($\vec{p}_j$).

Notice that the standard account of parameter setting requires that the learner be able to identify triggers. That is, the learner must be able to tell when a structure in the PLD is well-formed according to the current grammar. For any UG($\vec{p}_i$) that occurs, the non-well-formedness of at least some triggers with respect to UG($\vec{p}_i$) must be decidable, and feasibly so. It is interesting to note that this assumes a connection between feasible computability and the learnability of the associated language. The computability of the set of well-formed structures in the whole language is *not* assumed (cf. Chomsky 1980, p. 121; 1981a, pp. 11–13); only the feasible computability of the grammaticality of appropriate triggers in the PLD of any actually occurring language is required.

The "actually occurring" condition in this last remark is essential. Notice for example that there is no a priori reason to reject a theory that proposes parameters that can only be set by degree-1,000 data, any more than there is a reason to reject a theory that treats some degree-1,000 structures as well-formed. Structures of such high degree do not occur, and languages that could be learned only on the basis of evidence of such high degree would not occur. The explanation of their nonoccurrence is unproblematic. In any case, if this approach to the learning problem is correct, the feasible computability property must hold with regard to data and languages of the kind actually found. Lightfoot does not undertake to quell any doubts about whether this property does hold for the cases that he discusses, but it is plausible that it does given the small size of the structures involved.[1]

Another thing to notice about Lightfoot's account of parameter setting is how it contrasts with the situation in which linguistic structures are "filtered" out of the set of PLD. Lightfoot uses a reconstruction of the change from Middle English to Modern English to illustrate how some data that the learner might hear will simply be filtered out as ungrammatical if they cannot be squared with the UG given its previously set parameters. Thus the triggers will often be less than the total linguistic experience of the child. The idea implicit in this story is apparently the following: *A structure is considered deviant and not part of the trigger if it cannot be reconciled with UG given previously set parameters.* In Lightfoot's example, the learner is forced to regard some structures with their usual meanings as deviant, prompting the adoption of new lexical meanings for some verbs. The point I want to make here is that the filtering situation is like the parameter adjustment situation in an important respect. In both cases, the learner encounters a structure that is not well-formed given the learner's current grammar. However, a trigger for a new parameter setting must differ from a structure that is filtered out in at least the following respect: The trigger can be recognized to be well formed by an adjustment of the parameters to more marked settings.

Now let's look more carefully at the question of what counts as a trigger. Lightfoot says "the trigger consists of a haphazard set of utterances . . . utterances of a type that any child hears frequently . . . the trigger consists only of simple, unembedded material." The suggestion is, apparently, that the learner learns from *structures*. The exact nature of the structures is not explicitly specified, in spite of their importance to the argument. We can get a better idea of what Lightfoot is assuming, however, by considering the examples of parameter setting discussed in the target article. In each case, a simple structure is presented as an example that can motivate the adjustment of a parameter to a marked setting. The examples include the setting of the bounding node parameter to a marked value in French and Italian; the setting of a "Dutch government" parameter to a marked value in English; the setting of an AGR parameter to a marked value in Chinese. In each of these cases the simple structure that Lightfoot presents is an S-structure with traces (or something very similar). Lightfoot's examples suggest that triggers are completely specified S-structures containing traces of movement, S-structures that are not allowed by the current grammar of the learner. In the setting of the bounding node parameter in French, for example, Lightfoot presents the following example as evidence that a simple structure suffices to force the learner to a marked setting:[2]

Combien$_i$[$_S$ as-tu vu [$_{NP}$ e$_i$ [$_{N'}$ de personnes]]?

This structure would violate Subjacency if e$_i$ is a trace of combien$_i$ and if S were a bounding node (in addition to NP), and so a learner who could detect this situation would be forced to the more marked assumption that S is not a bounding node.

This picture of parameter setting is a little peculiar, since it appears that the learner who has not yet learned the grammar of French must already be able to determine the S-structures defined by the correct grammar. We face a "bootstrapping" problem here. There may be a way out, but the problem is serious and important, and bears directly on the plausibility of Lightfoot's main argument. Suppose, for example, that the PLD contains structures but not those that mark movement relations with traces, as in Morgan (1986). Then the mere presentation of a degree-0 S-structure that cannot be accommodated by an unmarked parameter setting will not suffice to show that simple PLD can force the adjustment of the parameter, since the S-structure itself is not available to the learner. Obviously, Lightfoot wants to show that *simple structures available to the learner* suffice to determine the appropriate parameter settings. As Morgan says, "for learnability theories that do include a source of structural information in input, one must ask whether this source is plausible" (Morgan 1986, p. 47). Lacking any account of what is really available to the learner, Lightfoot's argument is incomplete. Obviously, if degree-0 S-structures would suffice – but we agree that the child does not have access to such data – then the relevant degree-0 learnability hypothesis is not supported. The problem of figuring out what data a learner uses at early stages is serious and is of the center of some current research, but Lightfoot neglects the issue.

***2. Triggers, ungrammatical and marked constructions.*** A little reflection suffices to show that a learner given PLD containing S-structures who conforms to what I called the "standard" account of parameter setting would still be unable to learn English. There are at least two independent considerations that show that the question of what counts as a trigger must be much more complex. In the first place, learners of English are exposed to data that can be accommodated by setting parameters to values that are more marked than they are for English. Even if some children hear primarily grammatical sentences (cf., e.g.,

Newport 1977), many learners hear ungrammatical sentences without thereby being prompted to overgeneralize by setting a parameter to a more marked setting. Consider again the setting of the bounding node parameter. It is widely recognized that some violations of Subjacency are not as severe as others; they could well occur in the environment of a language learner. The following structures, for example, violate Subjacency but are perfectly comprehensible and could occur in a learner's environment.[3]

He is the person who [$_s$ they left before [$_{NP}$ speaking to t]].
He is the person of whom [$_S$[$_{NP}$ pictures t]] are on the table.
What [$_S$ did you wonder how [$_S$ John fixed t]]?

The appropriate analyses of these strings are also controversial, but the point is not. It is just that some ungrammatical strings that occur in a learner's environment could be made grammatical by adjusting the parameters, and yet the learner does not make the adjustment. So the question is: What special property must a structure have for it to be recognized as a trigger rather than as an ungrammatical string that could be made grammatical with a more marked setting of some parameter? To show that degree-0 data suffice, we must show not only that they would force an adjustment if they were in the trigger, but also that they have whatever other property is required to distinguish them from a merely deviant, ungrammatical structure.

The point can be put in another way. Every linguist has noted that there seem to be various marked constructions in human languages, constructions that seem to violate predominant regularities. Indeed, these are to be expected. It is plausible that once we have properly determined the core properties of a language, we can learn various kinds of special constructions that do not fall into the regular patterns dictated by UG. So again, the question is, how can the learner distinguish such cases from those that should prompt the adjustment of a parameter? The difference from the previous case is that in at least some such cases, linguists may be prepared to countenance the structures as grammatical even though they must be exceptional and peripheral relative to the core.[4] The standard account of parameter setting needs to be revised in order to provide real support for particular claims about the complexity of the primary linguistic data required for human language acquisition.

NOTES

1. See Stabler (forthcoming) for a formalization of similar problems in the framework of Chomsky (1986a). It turns out that even the most naive and straightforward approach to establishing the well-formedness of a structure is typically feasible, but no general decidability result has yet been established.

2. As noted above, the analyses and constraints on which Lightfoot's arguments are based are controversial. The controversy about Subjacency is well known. For example, Koster (1986) has argued against it, and Chomsky (1986a) has moved to treating it as a graded effect. In the case of the particular *structure* displayed here, Lightfoot refers to Kayne (1981) for support of the analysis, but a careful reading shows that Kayne does not take the empty element inside the VP to be a trace. Yet it is crucial to Lightfoot's account that the empty element be the trace of the movement of the quantifier, a movement subject to Subjacency. Kayne (1981) does not pursue the matter in any detail, but refers to his earlier work, Kayne (1975, pp. 29ff), in which he suggests that the QP's are base generated in preverbal position in these structures. Noting that the point is controversial, he also refers to other earlier work on the topic. In more recent work, Obenauer (1984, p. 157) mentions the advantages of Kayne's (1975, pp. 29ff) analysis, but does not pursue the question that is critical for Lightfoot, viz., whether the empty category is the trace of a movement subject to Subjacency. Koster (1986, pp. 5–52) argues, of course, for a nontransformational analysis of this structure. The movement analysis that Lightfoot needs is not entirely rejected, though: see Guéron (1981, p. 132, n. 44), Safir (1985, p. 317, n. 39), and Stabler (forthcoming).

3. See, for example, Chomsky, 1986a, ch. 6–7 for a discussion of these examples. I do not know of any careful studies of common syntactic errors. Some interesting errors that have occurred in literature and in the publicized remarks of celebrities are collected in Creider (1987, Appendix 3).

4. This point was made by Grimshaw (1987).

## Observing obsolescence

Nigel Vincent

*Department of Linguistics, University of Manchester, Manchester M13 9PL, England*

An interesting, important, and largely original question that is raised by Lightfoot's target article is whether it is possible to establish a connection between that subset of the language data available to the child, which nonetheless does not constitute part of the trigger experience, and the phenomenon of obsolescence whereby structures are lost from the historically evolving language. The attempt to explore such a link is surely to be welcomed. In the present commentary I wish in a constructive spirit to point out some difficulties, both logical and methodological, with the enterprise as Lightfoot conceives it.

Let us begin by considering the logic of the question in a little more detail. Assume a grammatical construction C, which at a given period in a language's history is regularly and widely attested, is in harmony with other structural properties of the language, and therefore gives rise to data which are sufficiently robust for the language learner to make use of. Assume further that this structure is, in Lightfoot's sense, degree-0 learnable.[1] In such circumstances there is every reason to believe that C will form part of the trigger experience. What then could make C obsolescent? Otherwise put, what could make C fail to be part of a new generation's trigger experience? Obviously not a change in UG, which is by definition constant across time and space. The only answer, assuming that trigger experience does not vary randomly from epoch to epoch, is that the relation of C to other aspects of the language's structure has changed so that C and the rest of the language – call it $\bar{C}$ (= the logical complement of C) – are incompatible in the sense that they trigger mutually exclusive parameter settings. Since C, *ex hypothesi*, has not changed, this change must arise in $\bar{C}$. Such is the logic of the account Lightfoot offers of the notorious case of *like*.

So far so good, but notice now that a further question arises. If an independent change leads to conflicting patterns in the data to which the language learner is exposed, what determines which pattern is selected? Lightfoot's general argument quite reasonably entails that learners are able to discard or ignore much that is irrelevant, including even quite major intrusive factors – such as being brought up by non-native speakers of the language they eventually acquire – and are able instead to attend to or home in on the central triggering data which will ensure that they develop an appropriately set core grammar of the language in question. One is immediately prompted to ask, therefore: Why, in the scenario we have sketched, was the change in $\bar{C}$ allowed to happen? Would not the existence of C in the trigger experience have served to inhibit such a change? Why, in other words, does the learner not ignore the innovative pattern and treat it as part of the nontrigger background experience?

One possible answer, compatible with Lightfoot's overall argument, is that the innovative change occurs with respect to a part of $\bar{C}$ that is in some absolute sense more central than C. Thus, in the case in point, what renders the older interpretation of *him likes the queen* impossible is a change in basic word order. Suppose that we were to postulate some general principle whereby word order parameters take precedence over lexically determined patterns such as the *like* construction. Things would then fall into place rather neatly. We could say, in general terms, that obsolescence will come about when data relevant to an aspect of UG that would be determined later in the matura-

tional process are rendered ineffectual by virtue of changes in data for which the parameters are set early.[2]

What has still not been resolved in this scenarios is, of course, the question of how the original innovative change is brought about. If obsolescence is due to a structural "knock-on" effect, effectively a kind of syntactic push-chain (Martinet 1955), there will always have to be at least one initiating change motivated externally in terms of changes in the community that provide the data to which the learner is exposed. Moreover, such original, enabling changes cannot be explained in terms of the interaction of the genotype and the triggering experience, since that way lies an infinite regress. Rather these changes must be determined either by variations in adult, or at least postmaturational, usage, or by population shifts and other kinds of socially motivated factors. This may seem obvious, but it has an important methodological consequence. Externally motivated changes will make structures obsolescent directly, without recourse to acquisitionally induced effects. Therefore, when we look back over the history of languages to spot potential cases of obsolescence to use in constructing theories about the limits of trigger experience, we will have to be able to distinguish between what we might call direct (i.e. socially determined) obsolescence and indirect (i.e. psychologically determined) obsolescence. Only instances of the latter would be germane to Lightfoot's case.

This is not the place to embark on a full account of how instances of the appropriate kind of obsolescence may be detected, but some preliminary remarks may be in order. One criterion, adumbrated in Lightfoot's own earlier monograph on syntactic change (Lightfoot 1979), would be the relatively sudden disappearance from attested sources of a construction. Whereas other kinds of grammatical change, such as grammaticalization (cf. Hopper & Traugott, forthcoming), seem to be gradual and cumulative over relatively long timespans, any change involving an incompatibility between two structures should manifest itself soon after the potentially "threatening" structure had emerged. Lightfoot (1979) applies the same argument to the identification of cases of reanalysis. A further guide might be to look for instances of obsolescence linked to reanalysis. Note, however, that reanalysis is not always entailed. Lightfoot cites the example of *like*, and to this we might add *lose*, whose Old English etymon *losian* had the sense "be lost to" so that the subject of the sentence expressed the thing lost and the sufferer of the loss was in the dative case (Allen 1986, p. 384). On the other hand, many more verbs from the class in question simply disappear from the language (*forðunchen* "to displease," *misliken* "to cause/feel dislike," *ofðunken* "to cause/feel regret" – cf. Allen 1986, p. 396) or survive only in minor and archaic patterns (e.g. *what ails thee?*, *to rue the day*, *it behoves me to say*, etc.). A more complex case concerns the disappearance of *ð unchen* in the sense of "seem." This verb also involves a dative experiencer, as does modern English *seem* (e.g. *it seems to me that*. . .). In this instance the borrowed item *seem* has taken over the old pattern, and the modern reflex *think* now has an experiencer subject (cf. earlier *methinks*). (See Vincent 1987 for more details.) Here the initial cause is manifestly external, being due to borrowing, which is in turn a special case of language contact. The latter is an obvious and competing source of explanation for obsolescence and, once again, one that would have to be eliminated in any future instances to be cited in favour of Lightfoot's position.

NOTES

1. Incidentally, these are not unreasonable assumptions to make about the Old English sentence pattern *him likes the queen* that Lightfoot takes as his central example. Nonetheless, the example is not without its problems. My personal preference is for a lexicalist account of the kind given by Allen (1986) which, if accepted, would invalidate the *like* case as an instance of the logical possibility which Lightfoot discusses. Since it is the logic of the argument which concerns me here, I proceed on the assumption that Lightfoot's example is a valid one.

2. Compare here Lightfoot's observations (sect. 1, para. 5) about the nonhomogeneity of UG.

## Why degree-0?

Thomas Wasow

*Departments of Linguistics and Philosophy, Stanford University, Stanford, CA 94305*

**Electronic mail:** *wasow@russell.stanford.edu*

The hypothesis that natural languages are degree-0 learnable, in the usual sense of the term, is clearly untenable. Translated into nontechnical terms, it amounts roughly to the claim that children can learn languages without being exposed to any example sentences that contain subordinate clauses. Lightfoot argues for a weaker position, admitting the need for subordinate clauses in the input data, but claiming that only the "front of an embedded clause" is necessary for learnability. Since fronts of embedded clauses occur only in sentences with backs of embedded clauses, it follows that embedded clauses are needed in the input data. Lightfoot's hypothesis, then, appears to be that children can attain grammatical competence without attending to the backs of the subordinate clauses in the sentences they hear. But this buys little, if anything, in accounting for language acquisition. Surely the child is trying to understand what is being communicated, not just to learn grammar; and ignoring the backs of embedded clauses will obviously make it very hard to understand what the speaker intends. Thus, once it is conceded that embedded clauses are needed in the input data, it is exceedingly odd to assume that children ignore all but the beginnings of those clauses.

Lightfoot argues for degree-0 learnability in his sense by showing how a few apparent counterexamples might be handled within one contemporary theory of grammar. But the examples he gives are by no means indicative of the number and variety of prima facie counterexamples available in the languages of the world. It is as though someone were to argue that a photograph, together with general principles of perspective, would always suffice to permit computation of the distances between the objects depicted. Although instances do exist where such computations are possible, they do not establish the general claim – which is, of course, false. In the space allotted to me, I will give a very superficial listing of a few simple counterexamples to the degree-0 learnability hypothesis, concentrating on ones that are inconsistent even with Lightfoot's expanded definition. While it may perhaps be possible to construct accounts for some of them, my list should at least establish where the burden of argumentation falls.

(i). Lightfoot himself (note 14) alludes to a well-known phenomenon that only occurs in embedded clauses: In standard German, the finite verb must be the last word in a subordinate clause, a position in which it never appears in main clauses. This is exemplified in (1).

(1) a. Ich weiss nicht, ob ich gewonnen habe.
I know not whether I won have
b. *Ich weiss nicht, ob ich habe gewonnen.
c. *Ich gewonnen habe.
d. Ich habe gewonnen.

It is hard to see how this contrast in word order could be mastered without exposure to the backs of subordinate clauses, since that is where the verbs in question occur. Moreover, there is variation across the Germanic languages with respect to this phenomenon, suggesting that it is unlikely to be predictable on the basis of universal principles.

(ii). One component of learning a language is learning the words in it. Many languages have words that appear only in embedded sentences, for example, English "if" and "lest."

Although these words appear at the beginnings of their clauses, there are forms of words that only appear in other positions in embedded clauses. An example from English is the subjunctive "were," exemplified in (2).

2. a. If I were the boss, I would fire the whole bunch.
   b. *I were the boss.

The following passage from Sells (1987, p. 445) describes another kind of word that occurs only in subordinate clauses:

> A . . . "logophoric" pronoun has a distribution distinct from that of other pronouns. . . . Roughly, the antecedent of the logophoric pronoun must be the one "whose speech, thoughts, feelings, or general state of consciousness are reported" (Clements (1975, p. 141)); hence logophoric pronouns appear predominantly within sentential arguments of predicates of communcation and experience.

Hyman and Comrie (1981) report that in the African language Gokana, logophoricity is marked on verbs, not on pronouns, and the verb in Gokana follows the subject; hence, logophoricity is not indicated at the front of the subordinate clauses. So, the logophoric forms of Gokana verbs and the English subjunctive "were" are words that never occur in degree-0 data. But then degree-0 learnability would entail that such words would not be learned.

(iii) Andersson and Dahl (1974, p. 453) describe a phenomenon in Swedish that only occurs in subordinate clauses: Forms of the cognate to the auxiliary "have" can be omitted in embedded clauses, where it would be required in the corresponding main clause. Thus, in (3a), if "hade" is omitted, the sentence becomes ungrammatical. It can, however, be embedded with or without "hade," as (3b) illustrates.

(3) a. a book which, I don't know who$_j$ to ask e$_j$ where to publish e$_i$
   b. *a man who$_i$ I don't know what$_j$ to ask e$_i$ where to publish e$_j$
   c. *a book which I don't know who I could ask where I should publish

A similar phenomenon exists in some dialects of English. A reduced form of the auxiliary "have" (represented here as "of") can appear in some embedded environments, where it would be impossible in the corresponding main clause. This is illustrated in (4).

(4) a. If I hadn't of waited, I would've won.
   b. *I hadn't of waited.

(I am indebted to Charles Fillmore for pointing this contrast out to me.) Evidently, the distribution of the Swedish "hade" and the reduced auxiliary illustrated in (4) could not be learned without access to degree-1 data.

(iv) Keenan and Comrie (1977) noted that languages differ according to what functions the relative pronoun can serve within the relative clause. For example, in Malagasy a relative pronoun must be the subject of the relative clause, in Welsh it can be either the subject or the direct object, and in English it can have any of a number of functions in its clause. None of the substantial literature dealing with this variation claims that it can be predicted from facts about sentences without any relative clauses; nor do the fronts of the relative clauses provide the necessary information. But this is precisely what would be required to maintain degree-0 learnability.

Such examples could be multiplied many times over. The traditional distinction between main clauses and subordinate clauses is based on the observation of generations of grammarians that there are systematic differences between them. Many of those differences are manifest at the front of subordinate clauses, but many others are not. The distinction is so deeply ingrained that at least one linguist (Williams 1973) published an article whose point was to argue that there was one language that did not differentiate main and subordinate clauses structurally! Given the pervasiveness of differences between main and subordinate clauses, together with the fact that those differences are not uniform across languages, the obvious hypothesis to adopt regarding learnability is that degree-1 data are necessary. It also seems plausible that degree-1 data are sufficient, since there are very few cases of putative differences between singly and doubly embedded clauses. Moreover, given Lightfoot's peculiar redefinition of degree-0 learnability, degree-1 learnability doesn't demand that the data available to the child be any richer than is required for Lightfoot's hypothesis. All that is required is that children learn from the data available. There is ample evidence that they do this, so why would one event want to pursue the degree-0 learnability hypothesis?

ACKNOWLEDGMENT

I am grateful to the following people for helpful advice in the preparation of this note: Judith Aissen, Chuck Fillmore, David Perlmutter, Stanley Peters, Geoffrey Pullum, Ivan Sag, and Peter Sells. None, of course, bears any blame for any errors of fact or judgment.

## Why degree-0?

Wendy Wilkins

*Department of English, Arizona State University, Temple, AZ 85287-0302*
**Electronic mail:** *atwkw@asuacad.bitnet*

In our work on learnability, Culicover and I (Culicover & Wilkins 1984) were motivated to restrict the learning model to degree-0 because our theory of grammar was based on specifically degree-0 phenomena: the assignment of thematic and grammatical relations. We were interested in showing how any error in the assignment of a grammatical relation would be detectable within the degree-0 domain in which grammatical relations are assigned. Nothing in Lightfoot's theory of grammar, however, specifically motivates degree-0 learning (as opposed to, say, degree-1). Although much can be learned from degree-0 input, Culicover and I were forced to move to "degree-0+," what Lightfoot subsequently calls "main clauses (. . .) plus a little bit." In point of fact, however, it seems more likely that the appropriate characterization of the child's PLD is degree-1.

Before presenting the sort of evidence from which I would argue for degree-1 learning, let me say that as far as I can tell, Lightfoot's overall theory would be neither weakened nor strengthened by a move to degree-1. His expanded characterization of degree-0 coincides with his theoretical notion of binding Domain, but this does not necessarily constitute an argument for either the theory of grammar or the theory of learning. The local domain of movement as determined by the Subjacency condition coincides with degree-1, but this, again, would not necessarily argue for the particular learning model. Before a beginning discussion of degree-1 versus degree-0, it must be reiterated that what is at issue is not what sort of data are in the child's linguistic environment (children surely have access to sentences with embedded clauses) but, rather, what sort of data are contained in the PLD from which learning takes place.

Consider first Lightfoot's discussion of example (24). The grammaticality of this example "dictates that S cannot be a bounding node in addition to NP." But nothing in Lightfoot's proposed model assures that the learner will conclude that it is S that is not a bounding node. The learner might wrongly conclude that NP is not a bounding node. Recovery from this incorrect hypothesis, given only degree-0 input, could not be achieved. To avoid this problem, Lightfoot would have to claim that NP is universally a bounding node, not subject to parametric variation. Only further theoretically informed analyses of more languages will tell whether this is an empirically justifiable result.

But consider a related issue with respect to bounding nodes. In French, the crucial datum for choosing S or S′ involves the extraction from NP of a pre-N′ constituent (example 24b). In Italian, it involves the extraction of the clitic *ne* from the subject of an embedded small clause (example 45). In Spanish, there is no movement similar to either the French or the Italian. And yet, as Torrego (1984) demonstrates, Spanish has S′ as the bounding node. Unless some appropriate extraction from NP in Spanish can be found (none has been proposed to date), this is a serious problem for the degree-0 learning of the bounding nodes for Subjacency.

Consider next a language-particular issue that is not directly related to the matter of parameter setting. It is generally assumed, as Lightfoot points out via the reference to Ross, and especially since Emonds's (1976) elaboration of the Structure Preserving Hypothesis, that there are no rules that apply only in embedded domains. It is interesting to recall, in this light, a certain problematic construction in Navajo. Kaufman, in her 1975 MIT dissertation (published later as Schauber 1979), discusses a certain rightward clitic movement which she argues is unbounded. Chomsky (1975) recognizes this to be problematic both for Ross (1967) and for his own Subjacency condition. It is of interest here not specifically because of the purportedly unbounded nature of the movement, but because the clitic can "attach only to an embedded verb" (Schauber 1979, p. 52). In other words, the movement rule functions (unboundedly) only within embedded domains. Compare (1) and (2).

(1) Jáanadeesb̨as nízinígóó Mary bił bééhózin
J. 1.F.drive 3.want.COMP*to* M. 3.with 3.be known
"Mary knows where John wants to drive to."

(2) *Jáan naagháhígíí shił bééhózind̨é.é
J. 3.come.COMP 1.with 3.be known.*from*
"I know where John comes from."

(Many thanks to Geraldine Keams for corroborating the Navajo grammaticality judgments.)

For (1), Schauber argues that the clitic originates on the most embedded verb ("drive") and then attaches to the verb glossed as "want" (although in more complex sentences it need not move to the highest embedded verb). The ungrammaticality of (2) (Schauber's [36]) is due to the movement of the clitic from the embedded verb to the matrix. Although it might be necessary to question whether Navajo has quite the clause structure that Schauber assumes (Jelinek, personal communication), her data would seem to require close scrutiny by anyone who wishes to limit the PLD to degree-0. How could the distribution of some element be learned if it can move onto various embedded verbs but never into the matrix clause?

Finally, consider cases where the characteristics of an embedded structure are determined by aspects of a higher clause. In Spanish, any verb embedded under *es necesario* must be in the subjunctive:

(3) Es necesario que venga / *viene.
is necessary that 3.come.Subj. / *3.come.Ind.
"It is necessary for him/her to come."

It is easy to find examples in many languages where some aspect of an embedded verbal complex depends on some characteristic of the matrix. For Spanish, the general use of the subjunctive could be learned from main clauses. However, the phenomenon illustrated in (3) is essentially degree-1. What must be learned is the relationship between the matrix and the embedded clause. For such cases, Lightfoot might redefine the binding Domain so as to include AGR with "dependent tense," or perhaps he could represent tense or AGR in some way in the embedded Comp. I would suggest that a more adequate alternative would be to simply recognize degree-1 data as part of the PLD.

Lightfoot has developed a learning model specifically to address the setting of the parameters made available in UG. Learners, in addition to setting all the necessary parameters, must ultimately learn all the language-particular details as well (such as the distribution of clitics in Navajo or the subjunctive in Spanish). These details actually present the biggest challenge for the theory of learnability because principles of UG or markedness are unlikely to be directly relevant. For the idiosyncratic details of particular grammars, a heavier burden of learning is placed on the learner. If it turns out that data of greater-than-degree-0 must be used for the learnability of peripheral details, there is little reason for artificially limiting the PLD to degree-0 for parameter setting.

## Linguistic variation and learnability

Edwin Williams

*Department of Linguistics, University of Massachusetts, Amherst, MA 01002*

**Electronic mail:** *ewilliam@umass.bitnet*

It is worth emphasizing the difference between what Lightfoot has accomplished and what Wexler and Culicover (1980) accomplished: Wexler and Culicover proved a theorem about a general rule writing system, whereas Lightfoot has defended an empirical speculation about a system whose general properties are not known. So Lightfoot's result should not be seen as an "improvement" on Wexler and Culicover's theorem, but rather a different kind of contribution. At the time of Wexler & Culicover's work it was generally assumed that degree-1 should be attainable, but this eluded proof; degree-0 was clearly impossible, but Lightfoot's interpolation of a complexity between 0 and 1 (which he calls "0 plus a little") is an interesting compromise, grounded as it is in the linguistic notion of "domain."

What is known at this point is a good deal about particular parameters; however, the general questions, such as "what is the inventory of parameters," "what is the general character of the dimensions of syntactic variation," and "what is the relation between triggering data and the setting of parameters" are not sufficiently answered to permit theorem proving analogous to Wexler & Culicover's results for the general transformational model of earlier generative work.

The second of these questions may have no answer – that is, there may be no general properties of dimensions of variation. On the third question, Lightfoot (under example 24) correctly but mysteriously points out that there is no necessary connection between locality and complexity, but he is surely assuming some sort of "tame" and general answer to this question, otherwise the locality conditions would be of no use in guaranteeing simple crucial data in the way that he illustrates in his examples. For example, it is commonly assumed that the trigger for a parameter is "relevant" to the parameter in some way, and not arbitrarily related to it, though of course it could be otherwise.

An empirical speculation is defended by searching out counterexamples and defusing them, and this Lightfoot does for a number of clear and relevant cases. The problem can be put in the following extensional way: Can two languages differ, and their degree-0-plus structures not differ? To Lightfoot's cases I would add the following for his consideration:

1. The German verb placement rules: How is the child to learn that verb second does not apply in both root and embedded sentences, if the verbs of embedded sentences are not included in the domain of the matrix clause?

2. Special properties of relative clauses: Presumably relative clauses will not be a part of the matrix domain, so any properties they have had better be deducible from some properties of the 0-plus structures. But what property of the 0-plus structures will tell you, for example, whether or not relative pronouns can delete?

3. Long anaphors: Anderson (1982) reports that reflexive pronouns in Icelandic must have antecedents within the smallest indicative clause that includes the reflexive, which is a more liberal requirement than is found in English, where the antecedent must be in the smallest clause of any type. In order for the Icelandic child to know that his anaphor is of the liberal type, he would seem to need an example of an anaphor in nonsubject position of an embedded nonindicative clause unambiguously taking an antecedent outside that clause; but this would not be a 0-plus structure.

The notion of domain relevant for degree-0-plus learnability cannot be tied directly to the notion of domain of reflexive binding, for in some languages (e.g., Dutch, as shown in Koster 1984) there are several distinct reflexives, each with its own binding domain, and so the binding of reflexives does not determine a single notion of domain for the language.

Finally, I think Lightfoot gives the impression that syntacticians generally ignore the question of learnability. My impression of the recent literature in neo-comparative syntax is that there is a widely felt obligation to outline the learning path whenever a parameter is posited, indicating whether negative data are required and how complex the crucial data are. There is in fact at least one volume of papers on the topic (Roeper & Williams 1987).

# Author's Response

## Matching parameters to simple triggers

David Lightfoot

*Department of Linguistics, University of Maryland, College Park, MD 20742*

**1. Adjusting the rigging.** Writing a target article for *BBS* peer commentary is a bit like sailing a 12-foot laser during a tornado alert. There are sure to be high winds, perhaps strong enough, if a tornado does come, to make one abandon the boat and one's dignity. In advance, one can expect that some gusts will turn in on themselves and amount to nothing, some will send the boat in a new direction, some will catch it and speed it forward, and others may require some redesigning. It's nerve-racking but irresistible.

Like a laser, the idea I was sailing was simple and limited, although it depended on some complex engineering. Any selective theory of language acquisition supposes that a child endowed with certain properties acquires certain kinds of linguistic capacities when exposed to certain kinds of linguistic experiences. On this much everybody agrees. I claimed that certain current assumptions about the endowed properties and the eventual capacities seem to suggest that children acquiring their syntactic systems need access only to very simple experiences, in fact only to unembedded material if one defines "unembedded" in terms of binding Domains at the level of logical form (and that is where some engineering is relevant).

Why should anybody care? It has long been known that not everything that a child hears has a noticeable or long-term effect on the emergent mature capacity; some sifting is involved. Some of the sifting must surely be statistical, some is effected through the nature of the endowed properties, some results from the cumulative effect of acquisition and the fact that children may not make any sense of certain expressions at some stage of development. And now I am suggesting that there may be structural limitations. If a plausible case can be made (and clearly that will take more than one article in *BBS*), then there will be one more tool for evaluating analyses and linguists will question proposals which entail more complex triggering experiences. The same is true, of course, if we resort to a weaker claim, that the triggering experience may include structures with one level of embedding, as several of the commentators think is plausible. For there is a more general point of logic, better intoned from the minaret than shouted from a laser: Selective theorists should, in principle, be able to point to plausible triggering experiences for whatever grammatical analyses they put forward. Except in some enlightened corners, I do not see with **Williams** "a widely felt obligation to outline the learning path whenever a parameter is posited." That's why I rigged the laser and wrote the article.

The complexity of the trigger experience is an important but limited concern. Many aspects of language acquisition are immune to it. For familiar reasons, genetically determined factors clearly permit the acquisition of even the first systematic sounds (contrary to what **Harris** seems to believe), but I doubt that anybody would imagine that exotic or complex data are needed to establish the inventory and distribution of sound segments and morphemes, or the meanings of most words. Nevertheless, these topics have been of great interest to people studying the emergence of language in children, and quite rightly. Issues of complexity do arise with the acquisition of stress and intonation patterns, where different analyses sometimes make different claims about how much a child would need to hear (see Chomsky 1972); recently there has been some fascinating work on this by Dresher and Kaye (forthcoming), who provide a model of how a child sets the parameters of a certain stress system. But my concerns have been even more limited, dealing with some complexity issues in the syntactic domain. This disappointed commentators who would have preferred a different topic.

Consequently, I had nothing to say about how children acquire the sounds of their language or come to use *wawa* as a word and a noun with the meaning roughly of "water." This is a nontrivial process, as **McCawley** notes, but it happens, and several people have examined how it happens. Having established that *water* is a noun, children later acquire the constituent structure of *water from the bath*, if I am right, by setting the parameters in (6) (see target article) and projecting to NP accordingly via N′, yielding $_{NP}$[Spec $_{N'}$[$_{N'}$[ $_{N}$[water]] $_{PP}$[from the bath]]]; for an excellent recent discussion of this aspect of language acquisition, see Lebeaux (1988). In setting these particular parameters, children are operating with partially formed representations which include $_{N}$[water], $_{P}$[from], $_{Spec}$[the] and $_{N}$[bath]. They are not operating with "raw data" or mere words, as **McCawley, Morgan, Snow & Tomasello,** and **Stabler** correctly point out. And they are

certainly not "judging" *water from the bath* to be grammatical.

*Water from the bath* and similar expressions occur in the child's environment with an appropriate frequency (more on this later), and, given a partially formed grammar whereby *water* and *bath* are classified as nouns, a projection can be assigned conforming to the schema of (6) (discussed nicely by **Stabler**). Contrast this with some ill-formed expressions that a child might encounter for various reasons: *from the bath water*, uttered by a guest who speaks a head-final language, is not a sufficiently frequent type to have any effect, and *water uit het bad*, uttered by a Dutch house-guest, could not be analyzed by a partially formed grammar in which *uit* is not classified. Of course, if the Dutch guest stayed long enough and spoke often enough about bath water, then *uit*, *het*, and *bad* might come to be classified as P, Spec and N respectively with the relevant meanings, and the child might interpret the string correctly as a phrasal complement of *water* and even perhaps use such phrases – in which case, one would witness an instance of code-switching, a frequent phenomenon among children exposed to more than one language. If there is a great deal of this kind of mixture in the trigger experience, the child will attain some form of creole. This is an extreme case which reflects the quite general capacity (invoked here by **Kroch**) to operate with a heterogeneous grammar. There is much more to be said about this, but there is nothing mysterious about it and children at this parameter-setting stage are neither judging sentences to be grammatical (*pace* **Buckingham**) nor judging words to be English (*pace* **Harris**). No particular issues of grammaticality are raised by this process, as far as I can see, and, despite the claims of Bickerton (*BBS* 7, 1984) and others, the phenomenon of mixed trigger experiences has not yet revealed anything special about the nature of UG.

In the target article I emphasized that children are not "little linguists." If they are genetically endowed in the way I suggest, then they undertake little analysis in the usual sense of the term; exposure to a sufficiently robust structure sets a parameter automatically and without any conscious analysis or effort. I surely have no reason to believe **Snow & Tomasello** when they assert (without evidence) that language acquisition involves a "long, slow, laborious process" – slow compared to what other labors? If children are endowed with the results of our linguistic arguments, such as a properly constructed phrase structure template, then they do not conduct any analysis "which parallels suspiciously closely the overt arguments" of linguists (**Harris**). Our poverty-of-stimulus arguments are the scaffolding that analysts may kick away when hypotheses about UG seem firm enough to pursue at other levels of abstraction, but children no more recapitulate these arguments than they rehearse the poverty-of-stimulus reasoning of Gregor Mendel or Thomas Hunt Morgan about the genetic basis of their physiological development.

A number of commentators discussed the "metaphor" or "analogy" to the biological sciences, but there is no analogy; this *is* biology and we use reasoning parallel to that of other biologists working at comparable levels of abstraction. Arguments from the poverty of the stimulus are not a "codicil to the theory" (**A. Grimshaw**) but as essential to the enterprise as they were for Mendel: Mendel noted properties and generalizations applying to pea plants which went beyond what could have been determined by environmental factors, given certain assumptions. He therefore hypothesized that pea plants developed according to an internally prescribed program which was subject to certain computational principles (the laws of segregation and independent assortment); he had very little idea of how these principles were instantiated. Later work explored these and other generalizations at biochemical levels, yielding fruitful ideas about the behavior and chemical structure of genes, and much more. If the kind of work I have described continues to be successful, there will be another synthesis, fulfilling the rational hopes of **Lorch:** We shall understand how notions like NP and the binding theory translate into other levels of abstraction such as neurological terms, and how they are realized biochemically, in the same way that Mendel's "factors" or "gens" (sic) of pea color, position on the branch, and so forth, were eventually seen to correlate with certain biochemical notions – despite the incredulity of his contemporaries (see Lightfoot 1982 and Jenkins 1979).

It is conceivable, presumably, that biology might have evolved differently, in such a way that Mendel's line of thinking would not have been a precursor to biochemical investigations. It is likewise possible that biochemists will next week announce a breakthrough in our understanding of the biological basis of language; but there is surely nothing irrational in pursuing work at a level of abstraction parallel to that of Mendel, in the hope that we can refine crucial ideas in such a way that they can later be linked to neurological or biochemical notions. We have been learning a great deal about the content of UG by working at this level of abstraction, and this is likely to continue. The fact that there is no synthesis with neurology or biochemistry at this time does not warrant a principled objection to such a synthesis. It hardly seems rational to reject the biological notions of triggering and parameter-setting on the grounds that languages are not species-specific like birdsong (**Harris**), that parameters are not set irreversibly and instantaneously by single events (**Snow & Tomasello**), or even that parametric variation fails to account for "the idiosyncratic details of particular grammars" which go beyond parameter-setting (**Wilkins**), which is a distinction without a difference. It is clear that the computational system supporting the human language capacity has some unique properties; we aim to discover what they are and to enrich our understanding of human biology accordingly. For example, one of the distinctive properties of UG is its plasticity, the fact that it is compatible with many different mature states, depending on the environmental factors a child is exposed to. This is not unique; Brandon and Hornstein (1986) point to the plant arrowhead (*Saggitaria saggitifolia*), which puts out different leaves above and below water. However, this kind of "phenotypic plasticity" (Brandon and Hornstein's phrase) has not been studied extensively by biologists; the variation found in human languages promises to extend our understanding of this concept.

One element of understanding how phenotypic variation emerges will be to determine not only the complexity

of the data which may act as a trigger but also the required robustness. Robustness is presumably a function of salience and frequency. We can be sure with **Snow & Tomasello** that parameters are not always set by single events; that would make the child too "trigger happy" and inclined to draw long-term conclusions (a metaphor) from too little data, as **Grodzinsky** shows (unless there is an independent sifting mechanism based on something other than frequency). However, this is not a trivial matter and different parameters may require more triggering experience than others. Indeed, some parameters may in fact be set by single events; for example, some child might "learn" (another metaphor) the meaning of the words *add* and *giraffe* from one exposure, particularly if that single exposure focused much attention on the word and provided a simple definition. Even some structural parameters might require few triggering experiences; it might not require more than a few instances of *water from the bath*, *student of physics* and such like to fix a prepositional complement as following its nominal head. As noted, raw data do not act as triggers, so it will not be enough to count how many such phrases are heard by the child under investigation. This parameter can be set only when the child has a partial analysis whereby *water* is classified as a noun and *from the bath* as PP; one will want to know how much it takes from that stage before the relevant parameter is set.

So any counting has to involve significant analysis and has to presuppose a good account of the primitives involved in the prior analyses; for example, is it enough to know that *water* is a noun and *from the bath* a PP, or does the child also need to know that the PP is in a modifying relation? To answer those questions and thus to provide a basis for the relevant quantification requires a better understanding of parametric variation among languages than we have at present; this quantification is hence not a matter high on my current research agenda. Nonetheless, there are real questions here, as **Buckingham** and **Schlesinger** point out, and it certainly seems plausible to assume with **Baker** that not all degree-0 data are robust enough to act as part of the triggering experience and have long-term effects. My sense is that more grammatical analysis is needed before we can develop precise and useful ideas of what the relevant robustness consists of. (Unlike Schlesinger, I see no reason to suppose that a child sensitive only to robust data would ipso facto and with no help from UG know the true extent of the partial generalizations which permeate natural languages.)

This serves to re-emphasize my point about the trading relation between ideas about UG, particular grammars and triggers, nicely discussed by **Cinque.** Notions about all three entities are theory-dependent and involve a good deal of analysis, as noted by several commentators; it is emphatically not the case that the triggering experience has some privileged accessibility to investigation, because it consists of partial analyses and few analyses are settled unquestionably and immutably. No useful claims can be made about the trigger without working assumptions about UG and particular grammars. This of course does not mean that one must first "settle," for example, the role of the Subjacency condition in accounting for the data in my (21), before raising issues about the possible triggering experiences, as **Buckingham** fears. Unlike **Stabler,** I do not see this as a problem, but just a normal situation occurring as theories are enriched: we build hypotheses on hypotheses, and most hypotheses are subject to revision and refinement as we go along.

**2. Tacking: Research strategies.** A sailor is always ready to abandon a tack when a wind makes it inappropriate, and some commentators advocated different tacks to get the laser to its target, different research strategies to understand language acquisition. Perhaps **Harris** was one such, but he worked up such a lather in Hong Kong and flailed at so many straw men that I was unsure whether he saw no laser or no lake.

**A. Grimshaw,** a sociologist, sees that linguistic arguments from the poverty of the stimulus are based on clear, impressive, and readily available data, and that hypotheses are therefore eminently testable and refinable (if he or others need more convincing on the uselessness of negative data, **Grimshaw & Pinker** will provide it forcefully). As a result, I am surprised that he is discomfitted by assumptions I took to need no discussion or demonstration: that a lot can be learned from constructed examples like *water from the bath*, *student of physics*, and so forth, regardless of who might have uttered such phrases under what conditions; that children sometimes resist correction; that certain kinds of rarely heard expressions may not be part of the triggering experience nor have long-term consequences. A. Grimshaw wants a research program with more attention to the social constraints on language, particularly when dealing with language change, but he does not show how this will cast light on the questions I was dealing with. Clearly, social factors account for why certain historical changes take place. Indeed, much of historical linguistics can be seen as a kind of population genetics, but not all of it. I'm not sure how Grimshaw thinks my point about obsolescence and triggering experiences would have been altered by considering social variables. It may well be that social factors will influence what I have called the robustness of a structure, but even that remains undemonstrated.

**McCawley** and **Schlesinger** are pursuing a research strategy which takes conceptual/logical categories as fundamental, but the differences may not be as great as they suggest. Pursuing a strategy which emanates from the generative semantics program, McCawley thinks that children have a genetically determined scheme of *conceptual* structure functioning in terms of logical categories (like predicate, argument, quantifier) that they can impose on expressions. Some of these notions may be necessary but they are by no means sufficient. My note 1 referred elliptically to the semantic distinction between *student from New York* and *student of physics*. If that analysis is correct, children eventually assign the structure $_{N'}[_{N'}[_{N}[\text{water}]]\ _{PP}[\text{from the bath}]]$ to *water from the bath*, and not $_{N'}[_{N}[\text{water}]\ _{PP}[\text{from the bath}]]$, although the PS rules of (2.a) make both structures available. This accounts for the fact that the phrase reflects two semantic "properties" (in the terminology of Baker 1978). It is entirely possible that this knowledge is part of the trigger for the syntactic structure; in that case, the trigger is not exclusively syntactic. In other words, the input to later stages of language acquisition includes not only a partial phonological representation of the expression but also a partial representation of the meaning; so the child needs a rough idea of the meaning of an expression as a syntactic

representation is assigned. This does not entail that Spec does not exist, that S must be subject to the PS template of (6), and least of all that some sort of conceptual structure suffices for the linguistic genotype.

The last is evidently what **Schlesinger** expects to find. He objects to the "special-purpose solutions" of generative grammar on the grounds that they cannot account for various selectional violations. He offers a generalization that subjects are viewed as either agents (his 1–3) or having responsibility (4–5). This is false: *The tent* in (4.a) and *this shirt* in (5.a) are in no way agentive or "responsible," and it begs the question of why *the car* is agentive in (1.a) but not *the wind* in (1.b). Even if the generalization were true, one would need to ask whether it was underdetermined by childhood experience. If it were generally valid, then it could be "learned" inductively by a child because all subjects would exhibit it and there would be no poverty-of-stimulus problem. If it is a partial generalization, then genotypical principles are likely to be involved. According to Chomsky (1965) children learn that the subject of certain verbs tends to be agentive and that this learning is guided by general principles of UG which delimit the form of the lexicon. Burzio (1981) provided analyses of ergative verbs (as in **Schlesinger**'s 4 and 5), which have the effect that they may not have agentive subjects. More generally, even if "innate constructs are of no avail in many other cases, which are explained . . . in terms of general cognitive factors," this would hardly entail that innate constructs are not needed for the kinds of cases I discussed. If innate constructs are relevant for certain things and not others then that would be the fact of the matter, whether "parsimonious" or not.

**Snow & Tomasello** study the speech of children and speech addressed to children. Much can be learned from this for my purposes; however, one hopes that they interpret children's speech more successfully than they interpret linguists' articles. Their response deals misleadingly with imputations and implications. They charge me with neglecting "virtually all of the empirical research on PLD," concluding that my concept of PLD is too narrow, because it ignores much of what children hear and do. But as **Wilkins** reiterates, what is at issue is not what data are in the child's environment, but which are the basis for "learning." Snow & Tomasello take "PLD" to be what children hear, whereas I used the term to refer to the subset relevant for language acquisition; therein lies the source of much misunderstanding.

Research on what children hear is not necessarily concerned with PLD in the sense I used the term; this is clear in the work of Wells (1981), who sought a comprehensive account of what children hear and understand at different stages. Such accounts are extremely useful insofar as they elucidate what children of various ages understand and how, but they are not necessarily studies of PLD (and nowhere did I suggest that Wells thought that PLD included everything children hear). The misunderstanding leads **Snow & Tomasello** to bark up a gum tree: They want to distinguish usable from nonusable speech, but usability is not the issue. Children can use and understand complex nursery rhymes, but that does not mean that complex structures are needed to establish the child's grammar. Nor does the fact that they sometimes imitate a novel and complex adult form entail that this form is part of the trigger (and nowhere did I suggest that children do not imitate). Such phenomena are important, but they are not what I was dealing with; nor, it seems, are they a good place to *start* to investigate the trading relation between UG and the trigger experience. A grammatical analysis must be provided for what the child understands, and that analysis will depend on details of UG. Where observational work on children has successfully been brought to bear on the trading relation under discussion here, it has customarily involved more grammatical analysis than one finds in the work of Snow and her associates.

These are issues of research strategy, however, and different strategies typically have different levels of success at different times. A strategy which has not been notably successful recently may bring unexpected results next week; that's one of the elements that makes scientific work and sailing exciting. Certainly it would be wrongheaded to believe that adherents of one approach have some monopoly on the human genotype, as **McCawley** points out.

**3. Scope and shape of UG.** There are many ideas on the range and nature of the linguistic genotype. **Schlesinger** and **O'Grady** put forward a Piagetian form of "general nativism" with no principles specific to the language faculty. Schlesinger admits pragmatic and semantic primitives, but does not address the poverty-of-stimulus arguments for more specific notions. The one argument he offers does not lead to his conclusion, as we have seen. O'Grady, however, takes my arguments head-on and postulates propositional, computational, and hypothesis-formation modules as part of the genotype, all couched in terms which range over objects other than language, presumably. The propositional module includes a "semantic coherence requirement," whereby each referring expression must bear a thematic role. It is not shown how such ideas are relevant to nonlinguistic domains, but O'Grady crafts an analysis of the interpretation of *one*. The idea is that *one* has no lexical content, and therefore cannot assign a thematic role; consequently, one does not find **the one of physics*, alluding to a student. (**Neale** is right to object to couching any anaphora rule in terms of reference; the problem is that anaphora has taken on a technical sense which makes it inappropriate also.) The reason is that *physics* lacks a thematic role, violating the semantic coherence requirement. Contrast this with *the one from New York*, where *from* assigns a thematic role to *New York*. However, this generalization cannot be correct. If *one* has no lexical content, that would be true *a fortiori* of a phonetically empty noun; nonetheless, (1.a) seems more acceptable than **the student of physics is older than the one of chemistry*, and (1.b) certainly is (here "e" indicates the phonetically empty noun, standing for *students* and *pictures* respectively). One wonders where *physics* in (1.a) and *Jim* in (1.b) get their thematic roles from. Furthermore, the well-formedness of (1.c) suggests that O'Grady's account of *one* itself requires amendment.

(1) a. ?I met two students of geology and three e of physics
b. I saw two pictures of Kim and three e of Jim
c. I saw the one of Tim

Presumably this is the kind of genotype that **McCawley** has in mind, although I'm unsure whether he would call

his logical categories not specifically linguistic. It might also be what drives **A. Grimshaw** to urge linguists to consider extralinguistic concepts like parsimony and efficiency as the basis of UG. **O'Grady** sees clearly that the challenge will be to state these general ideas specifically enough to solve linguistic problems. He asks for more time, which is fair enough, but he must acknowledge that this program has a long and so far not very successful history.

**Haider,** in his guise as *advocatus diaboli,* presents UG somewhat differently. Rather than the environment selecting values for the parameters of UG, UG triggers the child to look for certain things in the PLD. I see no substantive issue here and certainly not that the former mode reduces the role of "cognition." Nor do I see any incompatibility with the fact (also pointed out by **Snow & Tomasello**) that children may adopt intermediate analyses which do not match those of their adult models; a two-year-old might interpret its linguistic environment in such a way as to set a parameter one way, but later, with greater maturity and more parameters set, interpret things differently and set the parameter differently.

This is a possibility in principle and has sometimes been assumed to occur, but I am unaware of strong arguments for such a sequence of events. In fact, **Cinque** offers an interesting discussion of the distortions which can be provoked by the actual stages of language acquisition; he suggests an alternative, "instantaneous" account of the acquisition of the much-studied null subject option of Italian. This and Radford's (1988) analysis on which it is based, strike me as very plausible. He argues for delayed parameter-setting rather than resetting. Radford's argument that INFL and other nonlexical projections are not available in the early stages of acquisition may cast light on the acquisition of null subjects, which has been extensively discussed since Hyams (1983). If so, then we shall want to know whether elements of UG are subject to an internal clock of the kind that regulates the emergence of secondary sexual characteristics (Borer & Wexler 1987), or whether data triggering the existence of INFL are somehow less accessible than data triggering the existence of nouns and verbs. There have been attempts to show that certain aspects of UG only become available at relatively late stages in the acquisition process, but they have not been entirely convincing (see Lightfoot 1982, pp. 173–75, and Weinberg 1987). Such questions are on the research agenda, however, and are being explored actively.

**Koster** enters an objection to the use of the term "Universal Grammar" to characterize the initial state of the language learner because of the modular nature of language. I actually used "linguistic genotype" to encompass information which must be available to children independently of any experience with language, in order for the eventual mature linguistic capacities to emerge in the way they do. If one cuts some substantial corners, one can take the information to be genetically encoded in some fashion; those corners relate to the possibility of epigenetic developments and the like, which nobody will have the faintest idea how to approach until theories progress beyond their present rudimentary state (for some discussion, see Lightfoot 1982, p. 4). I adapted the usual terminology about UG, grammars, and PLD to refer to the relevant parts of the genotype, phenotype, and environment. We know very little now about how this information is organized and how it interacts with other aspects of the human genotype (see Lightfoot 1982, pp. 42–49). It may turn out that **O'Grady**'s general nativism is appropriate in certain areas and there may be principles which are crucial for language *and* for some other cognitive capacity. Perhaps, for example, the recursive capacity presupposed by the number system is also exploited for language, as has sometimes been suggested. Given the kind of parametrization that seems to be involved and the apparent plasticity of the genotype, it hardly seems likely that there will be nothing specific to language. Meanwhile, I see nothing inappropriate in using "UG" to refer to the relevant cluster of properties as we grope our way to some of their most salient features.

**4. Why degree-0?** Despite the alert, I took the laser out to focus attention on a neglected aspect of selective models of acquisition and to formulate a plausible but strong hypothesis: Parameters are set on exposure only to simple structures and nothing special has to be learned in order to use complex structures. I did not offer a proof of degree-0 learnability along the lines of Wexler and Culicover (1980), as **Williams** correctly emphasized. Nor did I offer a *principle* comparable to Subjacency or the Subset Principle, *pace* **Rizzi.** Nonetheless, if some such form of degree-0 learnability is plausible, one will seek to re-analyze proposed parameter settings which seem to need richer triggers. I discussed three such examples in the target article, and I shall consider some more in section 6 below.

**Haider** wondered what a degree-2 learnable parameter would look like, and **Morgan** offered an ingenious argument against degree-2 learnability based on statistical data about speech addressed to Roger Brown's (1973) children. Rizzi's (1982a) analysis of S and S′ as bounding nodes would suffice for Haider. Although the parameter does not exist in the form that Rizzi originally proposed (see **Rizzi**'s commentary and section 5 below) there was nothing inconceivable about it. Furthermore, within the framework of that analysis nothing a priori required a degree-0 trigger like my (45), and the sentences were readily comprehensible. For a child, of course, Rizzi's sentences (my 21) could be comprehended only when much of the grammar was in place, not in the early stages of acquisition. At the relevant stage it is possible that exposure to just a few such sentences would have sufficed to set the parameter. If Morgan is right, children can be expected to hear 21,000 sentences of degree-2 or greater complexity over a five-year period; from that pool Italian children would need to hear a few like my (21) at the appropriate stage, regardless of the ratio of tokens-to-types. As a result, Morgan's principled case against degree-2 learnability fails, unfortunately.

These matters are hypothetical if some degree-0 story is correct. If children set parameters only on the basis of unembedded Domains, then **Rizzi**'s parameter could not have been set by structures like my (21). Furthermore, a parameter could not be set by properties of the verb or direct object of an embedded clause (but see below). Several commentators regard this as implausible, including **Lasnik,** and I shall address some arguments to that effect in section 6. This is nonetheless an empirical issue for a degree-0 account and a strong hypothesis. (It does

not mean, of course, that children only *hear* the front of an embedded clause, as in **Wasow**'s caricature.) I am therefore puzzled by Lasnik's contention that the difference between degree-0 defined in terms of binding Domains and degree-1 defined in terms of clauses (in the fashion of Wexler & Culicover 1980) may be terminological. It is of course possible that a binding Domain may be defined parametrically in some language in such a way that my notion of degree-0 would match Wexler and Culicover's notion of degree-1 *in that language*, but that would not make the ideas generally equivalent.

Not only is degree-0 learnability an empirical matter, but it also does not follow in any strong sense from the nature of current ideas abut UG. Locality conditions in UG certainly facilitate degree-0 learnability and probably constitute a necessary condition, but they do not suffice. Perhaps this is what **Williams** means by a "tame" relation. Again, **Rizzi**'s analysis of bounding nodes, while not correct, cannot be ruled out a priori and illustrates how a grammar with only local processes may nonetheless require complex data to set some parameter. The commentaries of **Joshi** and **Wilkins** are important in this regard. Wilkins is right to say that nothing in my theory of UG specifically motivates degree-0 learning, and that this theory of UG would be neither weakened nor strengthened if degree-1 embeddings proved to be needed for learnability . . . except that this would free linguists to postulate parameter settings sensitive to embedded verbs and direct objects. I do not view this as a shortcoming; indeed Culicover and Wilkins (1984) used a theory of UG with very strict locality conditions; this may have motivated them to investigate degree-0 learnability, as she says, but it did not guarantee degree-0 learnability. Wilkins now takes degree-1 embeddings to be available. Joshi, on the other hand, offers a mathematical specification of degree-0 learnability by using a form of UG which allows a set of elementary structures and a single adjunction operation to yield recursive structures. The locality of the grammar is taken to guarantee the simplicity of the PLD, but something more is needed to show how the German child, for example, sets the verb as VP-final and allows it to occur to the left only in root clauses. It will be an interesting result if degree-0 learnability can be guaranteed by some form of UG, but I know of no such argument yet and consequently see no *principled* reason for it.

**5. Beating upwind: Empirical issues.** The commentaries have raised some intriguing empirical issues, but here I can only briefly mention some general ones which require elaboration or amendment in my account. **Buckingham, Clark, Grodzinsky, Lasnik, Morgan, Wilkins,** and **Williams** all raise an important and general issue: how to match triggers to parameters. As indicated in section 1 above, this involves questions of robustness; the child must not be, in Grodzinsky's phrase, "trigger happy," reacting too readily to everything that is heard. An appropriate notion of robustness would solve Grodzinsky's problem with *give* and *donate:* Structures like his (1.b) would be robust in the appropriate sense, whereas (3.a) might occur in isolation but would not be sufficiently robust (that much would be plausible; unlike Grodzinsky and Randall (1985), I do not hear a great difference between (3.a) and (3.b)).

Under current analyses, however, the issue also involves an apparent multiplicity of triggers for a given parameter; so **Wilkins** points out that French, Italian, and Spanish would each use different triggers to set S′ as a bounding node. **Buckingham** asks whether the existence of null subjects would be set in a given language by the presence of postverbal subjects, empty subjects or, we add, "uniform" inflectional marking on verbs (see **Morgan**) (but surely not by the existence of object-verb structures, which do not occur in Italian and are presumably not relevant). The issue is a general one even if these examples do not survive further analysis: There is probably no S/S′ option for bounding nodes (below) and null subjects are probably set by hearing null subjects (which in turn are licensed but not triggered by a suitable INFL).

**Clark** shows that structures which seem to trigger "Exceptional Case Marking" in English (whereby a verb governs and assigns case to the subject of an infinitival complement) may not trigger an equivalent process in Irish and Latin. The Irish child, on exposure to Clark's (2), must somehow determine that the subject receives its case from within its own clause (from an INFL adjoined to S, if Chung & McCloskey 1987 are right), whereas the English child, on hearing *he expected her to win*, determines that *her* receives case from the higher verb.

**Lasnik** raises an intriguing problem involving the matching of trigger to parameter. He points to a grammar with encliticization and procliticization processes and asks what would prevent a child from subsuming some particular phenomenon under the wrong process. His answer is either indirect access to negative data (and he provides a very plausible scenario) or degree-1 embeddings. The problem may prove to be a general one and other examples will be interesting, but there are at least two plausible tacks for Lasnik's case. One would make cliticization processes generally subject to a government condition (as suggested in Aoun & Lightfoot (1984)). Since *want* but not *win* governs the INFL *to* in a structure (2) (assuming INFL to be the head of S), *to* must *en*cliticize on to *want*.

(2) $_{V'}$[want $_{S}$[to $_{VP}$[win]]]

Meanwhile simple cases like *s'Kim here?* or *s'cold in here*, with the clitic form of *is* in sentence-initial position, would show that *is* may be adjoined to the left of the following item (thereby governed by *Kim*). There's much more to be said about this, but the approach looks promising. The second tack would be to say that Lasnik's primary datum (6.c) *who do you think's here?* is a degree-0 datum: If *is* manifests INFL, which, in turn, is the head of the embedded clause, then it has the upstairs clause as its binding Domain.

**Rizzi**'s rich commentary offers an alternative means of defining "unembedded": in terms of government domains. He takes parameters to relate to the properties of heads and proposes that all such parameters should be settable by inspecting the immediate environment of the head, that is, the government domain. To determine the properties of, say, verbs, it should be necessary to inspect only a maximally accessible verb, that is, the verb of a matrix clause. This is an attractive formulation which avoids a so far unmentioned problem with using binding Domains as the key notion: If Domains are defined in terms of accessible SUBJECTS, then the direct object of

an embedded clause has no accessible SUBJECT, hence no Domain, and therefore has the matrix clause as its Domain by default (see Aoun 1986). By this reasoning children should have access to embedded direct objects in order to set parameters, which is an unwelcome relaxation of our central hypothesis. Rizzi's formulation does not have this defect and is therefore empirically distinct. Then the question arises of whether this view of parameters is exhaustive; is there variation in ways which do not relate directly to the properties of heads, for example, in terms of how binding Domains are defined? If the head-related view is not exhaustive, then Rizzi's definition of degree-0 will need to be amended. Also, the deletability of complementizers is manifested only in embedded structures; an embedded Comp would not be accessible to Rizzi's child. This may also distinguish the two approaches, unless it can be shown that the deletability of an item in Comp is a reflex of properties of the governing verb.

**Kroch** and **Vincent** offer interesting discussion of the implication of degree-0 learnability for work on language change. Vincent takes up the matter of obsolescent structures and gives an account which conforms in general to my understanding. He correctly distinguishes externally and internally motivated changes, pointing to the necessity of external changes due to adult innovations, population shifts and the like. He then wonders how one could distinguish obsolescence due to "a structural 'knock-on' effect" from that which is socially determined. This question is quite general in diachronic work and there are criteria for drawing such distinctions, as he notes. However, what is intriguing about obsolescence is that it must generally be due to structural factors, since only under quite exceptional circumstances is it plausible to argue that people *stopped* saying certain things for stylistic reasons or because of external influences; usually a structure becomes obsolete as a consequence of some other change and is, in some sense, pushed out. It is for this reason that I have argued that obsolescent structures are particularly illuminating (Lightfoot, in press).

Changes involving obsolescence seem relatively abrupt, like other structural changes, whereas externally induced changes seem to be more gradual; I do not see gradualness as a general property of historical change. There has been a continuing debate on this point, and **Kroch**'s views may be somewhat different. The quantitative research of Kroch and his associates is among the most important recent work in historical linguistics. He shows that people may operate with heterogeneous grammars; his statistics show that changes spread gradually but in a nonetheless systematic fashion reflecting the clustering of phenomena that grammars require. However, although all changes spread through a speech community gradually, it is less clear whether they affect individuals with equal gradualness. This is an important topic, requiring more extensive discussion (Lightfoot, forthcoming).

I discussed the available bounding nodes for the Subjacency condition within the framework of Rizzi's 1982a analysis to illustrate the markedness conventions involved and to make some points about the nature of the trigger. Grimshaw (1986) has shown, however, that Rizzi's analysis is probably incorrect, as noted in the target article and by **Buckingham, Koster,** and **Rizzi.** Furthermore, if Rizzi is right in his speculation here that parametric variation is limited to the properties of heads, then his 1982 proposals would be excluded in principle and they cannot be taken as typical of anything. His speculation, if correct, would also undermine the usefulness of the distinction drawn by **Freidin & Quicoli** between rule- and principle-related parameters.

**6. Running downwind: Empirical challenges to degree-0.** Several commentators devised challenges for degree-0 learnability. Some can be analyzed away in such a fashion as to strengthen my notion and to provide a following wind to speed the laser along. However, the most dangerous time on the lake is when running before the wind. Other challenges are more mysterious and may eventually cause a gibe and a capsize. Let's first get up some speed.

Two issues were raised by several commentators: verb-second phenomena in Dutch and German and long-distance anaphors in the Scandinavian languages. The apparent problem for the Dutch child concerns how to attain from unembedded information the knowledge that verbs occur at the end of the sentence in embedded clause. A Dutch child endowed with a phrase structure schema like my (6) knows that heads occur adjacent to complements in D-structure, perhaps preceding or perhaps following them. Partially analyzed structures like (3) might lure the child to the view that Dutch has subject [verb object] at D-structure; but (4), where the verb is not adjacent to its direct object, forces the child to conclude that something has moved.

(3) $_{NP}$[vele mensen] $_{V}$[lezen] $_{NP}$[geen krant]
"many people read no newspaper"

(4) in Utrecht lezen vele mensen geen krant
"in Utrecht read many people no newspaper"

Various factors would eventually show that it was the verb which has moved upward rather than the subject *vele mensen* downward into the VP, or any of the other possibilities. And the presence of specifiers and separable prefixes in final position (5) would show that to be the source position for the verb.

(5) a. vele mensen lezen de krant niet/nooit
b. vele mensen bellen de docent op
"many people call the teacher up"

Space limitations preclude further discussion here, but see Lightfoot (forthcoming). The phrase structure schema is the crucial factor indicating movement; a more aggressive sailor might take out another boat and hone this into an argument for the "special nativism" that some commentators seek to avoid.[1]

**Rizzi** pointed to one approach to long-distance anaphors in Icelandic, following Pica (1987) and keying their syntactic properties to their morphological form. Sigurjonsdottir and Hyams (1988) offer another approach, which treats Icelandic *sig* as a bound variable analogous to *his* in *everybody loves his mother*, and not as an anaphor. If it were an anaphor, then the Subset Principle would predict that early grammars would allow only local binding (because that defines a smaller language than a grammar also allowing long-distance binding). However, they show that children control the complex constraints on long-distance binding at a relatively early age and that

both children and adults show a strong preference for long-distance readings where the anaphors are ambiguous. Bound variables, on the other hand, typically have long-distance dependencies (*everybody hopes that his mother is happy*). Chien and Wexler (1987) and Lee and Wexler (1987) report different developmental data for Chinese and Korean, so the solution does not carry over straightforwardly for those languages. However, Baker's account of Wilcoxon (1988) suggests that apparent embedded reflexives in Chinese may be instances of an "intensive adverb," which will change the problem significantly.[2]

Relative clauses posed problems for some commentators. **Williams** wondered how the deletability of relative pronouns would follow from degree-0 properties. The binding Domain for a Comp is the upstairs clause and therefore the properties of an embedded Comp are available to a degree-0 learner; these properties include the deletability of items in that position and the features of words like *if* and *whether*, which **Wasow** did not see how to treat.[3] The availability of an embedded Comp might also suffice for Wasow's problem with Keenan and Comrie's (1977) accessibility hierarchy; insofar as relative pronouns bear the necessary case markers, a child will observe whether a relative pronoun in Comp may be a subject (Malagasy) or subject/object (Welsh) or whatever. Relative pronouns do not always carry the necessary markers, however, and in those cases one must hope that Cinque's (1981) approach is fruitful and that the generalizations correctly predicted by Keenan and Comrie's accessibility hierarchy emerge from principles of UG interacting with degree-0 features of particular grammars. Clearly, degree-0 learnability, like any other hypothesis, will be refined or refuted by counter*analyses*, not by raw phenomena. **Baker,** on the other hand, raises a problem for which I now have nothing to suggest: the occurrence of resumptive pronouns in relative clauses in some languages, or, comparably, the occurrence of object traces in English infinitival complements.

The other challenges to degree-0 about which I have only questions and nothing useful to say are **Wilkins**'s material on Navaho clitics, which appear to attach only to embedded verbs, and **Haider**'s discussion of extraction from within embedded questions and relative clauses (where it is unclear what parameter is involved, let alone how it is set). Wilkins mentions the subjunctive mood of verbs under predicates like *necessario* in Spanish; this is presumably a form of complementation, marked in the Comp or the head of the clause or both. However, questions of degree-0 learnability do arise from the "obviation" properties of Romance subjunctives: The subject of a subjunctive complement must be disjoint in reference from the higher subject (6).

(6) a. je veux que Pierre/*j'aille à Paris
I want that Peter/I should go to Paris

Maybe the subjunctive has the effect of making the binding Domain for the embedded subject the next clause up, like an infinitive; or maybe a pragmatic solution is in order.

**7. Conclusion.** There are therefore several substantial puzzles for this kind of degree-0 hypothesis. All such puzzles will involve matters of detailed analysis at the level both UG and the particular grammar under investigation. As **Cinque** points out, this is both a strong and a desirable hypothesis. I have argued here that it is also plausible; nothing in the commentaries leads me to abandon the boat. It is for others to judge whether dignity has been abandoned. However, the commentaries have raised several valuable questions about the nature of the parameter-setting model. This is important. Many conceptual advantages are claimed for the parameter-setting model over the earlier grammar selection model, whose child formulated and tested hypotheses according to an evaluation metric. The notion of parameters has provided a vehicle and an incentive for more productive comparative work, with important consequences. However, alongside the real conceptual advantages, there are many ambiguities, unclarities, and difficulties for such a model. These commentaries have been particularly interesting for the questions raised about it. I have tried to provide partial answers where possible, but fundamental questions are still open. I am deeply grateful to all the commentators for their energetic attention, for helping to tidy up the design of the boat and for suggesting how it could be sailed better.

ACKNOWLEDGMENT

Preparation of this response was facilitated by a fellowship from the American Council of Learned Societies.

NOTES

1. If some such account were not available, one might maintain a version of degree-0 learnability by taking the embedded verb to manifest INFL (fairly uncontroversial) and taking INFL to be the head of the embedded clause, whose Domain is therefore unembedded.

2. Perhaps this would also change **Lasnik**'s problem with object reflexives in Chinese. However, notice in this case that the binding Domain for the object *ziji* in Lasnik's (8) is, by default, the matrix (because it has no accessible SUBJECT). In my discussion of **Rizzi**'s alternative formulation (section 5 above) I noted the availability of embedded objects as a difference between our accounts. If no better solution is found for Lasnik's problem, there will be a reason to reject Rizzi's more restrictive formulation.

3. **Wasow** also points to *were* in *if I were the boss, I would . . .*, which occurs only in conditional clauses. Such forms are probably archaisms, more taught than acquired through the usual parameter-setting process, and thus by no means uniform in the speech of English speakers. Again, however, *were* occupies INFL and may be the head of the clause; in that case it might be available to language learners. Wasow also points to "logophoric" pronouns, citing Sells (1987) to the effect that they occur "predominantly" in complement clauses. By a useful rhetorical move, predominantly occurring in complement clauses translates a few lines later into "never occur in degree-0 data." Latin certainly allows logophoric pronouns in main clauses which are understood to reflect what somebody said. Perhaps this is also possible in Gokana; in fact, Wasow's source, Hyman and Comrie (1981), provides such an example in the final footnote. I have not been able to check his material on the deletability of Swedish *hade*, but he relates it to a similar phenomenon in some dialects of English which allow a reduced form of *have* (written *of*) "in some embedded environments." However, those dialects allow the same reduced forms in matrix clauses: *I could of done it*.

# References

*Letters a and r appearing before authors' initials refer to target article and response respectively*

Abe, N. (1988) Polynomial learnability and locality of formal grammars *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics* (ACL), Buffalo, NY, June. [AKJ]

Adams, M. (1984) Multiple interrogation in Italian *Linguistic Review* 4:1–28. [aDL]

(1987) From Old French to the theory of pro-drop *Natural Language and Linguistic Theory* 5:1–32. [RF]

Allen, C. (1986) Reconsidering the history of *like*. *Journal of Linguistics* 22:375–409. [aDL, NV]

Anderson, A. R. & Belnap, N. D., Jr. (1975) *Entailment*, vol. I. Princeton University Press. [JDM]

Anderson, J. M. (1986) A note on Old English impersonals. *Journal of Linguistics* 22:167–77. [aDL]

Anderson, S. (1982) Types of dependency in anaphors. Manuscript, University of California at Los Angeles. [EW]

Anderson, S. R. (1986) The typology of anaphoric dependencies. Icelandic (and other) reflexives. In: *Topics in Scandinavian syntax*, ed. L. Hellan & K. Koch Christensen. Reidel. [HH, LR]

Andersson, A-B. & Dahl, O. (1974) Against the penthouse principle. *Linguistic Inquiry* 5:451–54. [TW]

Aoun, J. (1986) *Generalized binding*. Foris. [arDL]

Aoun, J. & Lightfoot, D. W. (1984) Government and contraction. *Linguistic Inquiry* 15:465–73. [rDL]

Baker, C. L. (1978) *Introduction to generative-transformational syntax*. Prentice-Hall. [arDL]

(1979) Syntactic theory and the projection problem. *Linguistic Inquiry* 10:533–81. [aDL]

(1982) Review article on Wexler & Culicover (1980) *Language* 58:413–21. [aDL]

Baker, C. L. & McCarthy, J. J. (1981) *The logical problem of language acquisition*. MIT Press. [JG]

Berwick, R. C. (1985) *The acquisition of syntactic knowledge*. MIT Press. [aDL, RC, LR]

Bickerton, D. (1984) The language bioprogram hypothesis. *Behavioral and Brain Sciences* 7:173–221. [arDL]

Bloom, L. (1970) *Language development: Form and function in emerging grammars*. MIT Press. [ADG]

Bohannon, J. N. & Stanowicz, L. (1988) The issue of negative evidence: Adult responses to children's language errors. *Developmental Psychology* 24:684–89. [JG, CES]

Borer, H. & Wexler, K. (1987) The maturation of syntax. In: *Parameter setting*, ed. T. Roeper & E. Williams. Reidel. [rDL GC]

Bowerman, M. (1982) Reorganization processes in lexical and syntactic development. In: *Language acquisition: The state of the art*, ed. E. Wanner & L. Gleitman. Cambridge University Press. [CES]

(1985) What shapes children's grammars. The crosslinguistic study of language acquisition. Vol. 2: *Theoretical issues*, ed. D. I. Slobin. Erlbaum. [ADG]

Braine, M. D. S. (1971) On two types of models of the internalization of grammars. In: *The ontogenesis of grammar: A theoretical symposium*, ed. D. I. Slobin. Academic Press. [IMS]

(1988) Modeling the acquisition of linguistic structure. In: *Categories and processes in language acquisition theory*, ed. Y. Levy, I. M. Schlesinger & M. D. S. Braine. Erlbaum. [IMS]

Brandon, R. & Hornstein, N. (1986) From icons to symbols: Some speculations on the origin of language. *Biology and Philosophy* 1(2):169–89. [rDL]

Bresnan, J. (1971) *Contraction and the transformational cycle*. Indiana University Linguistics Club. [HL]

Brown, R. (1973) *A first language: The early stages*. Harvard University Press. [rDL, ADG, JLM]

(1977) Introduction. *Talking to children. Language input and acquisition*, ed. C. E. Snow & C. A. Ferguson. Cambridge University Press. [ADG]

Brown, R. & Hanlon, C. (1970) Derivational complexity and order of acquisition in child speech. In: *Cognition and the development of language*, ed. J. R. Hayes. Wiley. [JG]

Burzio, L. (1981) *Italian syntax: A government-binding approach*. Reidel. [rDL]

Canale, M. (1978) Word order change in Old English: Base reanalysis in generative grammar. Unpublished doctoral dissertation, McGill University. [aDL]

Changeux, J.-P. (1980) Genetic determinism and epigenesis of the neuronal network: Is there a biological compromise between Chomsky and Piaget? In: *Language and learning*, ed. M. Piatelli-Palamarini. Routledge & Kegan Paul. [aDL]

(1983) *L'homme neuronal*. Fayard. [aDL]

Chien, Y.-C. & Wexler, K. (1987) A comparison between Chinese-speaking and English-speaking children's acquisition of reflexives and pronouns. Paper presented at the 12th annual Boston University Conference on Language Development. [rDL]

Chomsky, N. (1965) *Aspects of the theory of syntax*. MIT Press. [arDL, IMS, EPS]

(1970) Remarks on nominalization. In: *Readings in English transformational grammar*, ed. R. Jacobs & P. R. Rosenbaum. Ginn. (Page references to reprint in Chomsky 1972b.) [JDM]

(1972) *Language and mind*, enlarged edition. Harcourt, Brace & Jovanovich. [rDL]

(1972a) Some empirical issues in the theory of transformational grammar. In: *Goals of linguistic theory*, ed. S. Peters. Prentice-Hall. (Page references to reprint in Chomsky 1972b.) [JDM]

(1972b) *Studies on semantics in generative grammar*. Mouton. [JDM]

(1973) Conditions on transformations. In: *A festschrift for Morris Halle*, ed. S. Anderson & P. Kiparsky. Holt, Rinehart & Winston. [aDL]

(1975) *Reflections on language*. Pantheon. [GC, WW]

(1980) *Rules and representations*. Columbia University Press. [EPS]

(1981a) *Lectures on government and binding*. Foris. [aDL, RC, HH, HL, EPS]

(1981b) Principles and parameters in syntactic theory. In: *Explanation in linguistics: The logical problem of language acquisition*, ed. N. Hornstein & D. Lightfoot. Longman. [RF]

(1986a) *Barriers*. MIT Press. [LR, EPS]

(1986b) *Knowledge of language: Its nature, origin, and use*. Praeger. [JLM]

(1988) *Language and problems of knowledge*. MIT Press. [JLM]

Chung, S. & McCloskey, J. (1987) Government, barriers, and small clauses in modern Irish. *Linguistic Inquiry* 18:173–237. [rDL, RC]

Cinque, G. (1981) On Keenan and Comrie's primary relativization constraint. *Linguistic Inquiry* 12(2):293–308. [rDL]

Clark, R. (1977) What's the use of imitation? *Journal of Child Language* 4:341–49. [CES]

(1978) Some even simpler ways to learn to talk. In: *The development of communication*, ed. N. Waterson & C. Snow. Wiley. [CES]

Clements, G. N. (1975) The logophoric pronoun in Ewe: Its role in discourse. *Journal of West African Languages* 2:141–77. [TW]

Creider, C. (1987) Structural and pragmatic factors influencing the acceptability of sentences with extended dependencies in Norwegian. *Working papers in Linguistics* 4. University of Trondheim. [EPS]

Culicover, P. W. & Wilkins, W. K. (1984) *Locality in linguistic theory*. Academic Press. [arDL, WW]

Demetras, M. J., Post, K. N. & Snow, C. E. (1986) Feedback to first language learners: The role of repetitions and clarification questions. *Journal of Child Language* 13:275–92. [JG]

Dresher, B. E. & Kaye, J. (forthcoming) A computational learning model for metrical phonology. *Cognition*. [rDL]

Ellegård, A. (1953) *The auxiliary "do": The establishment and regulation of its use in English*. Almqvist and Wiksell. [AK]

Elmer, W. (1981) Diachronic grammar: *The history of Old and Middle English subjectless constructions*. Niemeyer. [aDL]

Emonds, J. (1976) *A transformational approach to English syntax*. Academic Press. [WW]

Epstein, S. D. (1982) Clitic rules and learnability. Unpublished manuscript, University of Connecticut. [HL]

Ervin-Tripp, S. & Mitchell-Kernan, C. (1977a) Introduction. *Child discourse*, ed. S. Ervin-Tripp & C. Mitchell-Kernan. Academic. [ADG]

eds. (1977b) *Child discourse*. Academic. [ADG]

Felix, S. (1987) *Cognition and language growth*. Foris. [GC]

Fillmore, L. W. (1979) Individual differences in second language acquisition. *Individual differences in language ability and language behavior*, ed. C. J. Fillmore, D. Kempler & W. S-Y. Wong. Academic. [ADG]

Fisher, O. & van der Leek, F. (1983) The demise of the Old English impersonal construction. *Journal of Linguistics* 19:337–68. [aDL]

Fontaine, C. (1985) Application de methodes quantitatives en diachronie: L'inversion du sujet en français. Thèse de Maîtrise. Université du Quebec à Montréal. [AK]

Geach, P. (1962) *Reference and generality*. Cornell University Press. [SN]

Gold, E. M. (1967) Language identification in the limit. *Information and Control* 10:447–74. [JG, JLM]

Greenberg, J. H. (1966) Some universals of grammar with particular reference to the order of meaningful elements. In: *Universals of language*, ed. J. H. Greenberg. MIT Press. [aDL]

Griffin, D. R. (1984) *Animal thinking*. Harvard University Press. [JDM]

Grimshaw, A. D. (1987) Disambiguating discourse: Member's skill and analyst's problem. *Social Psychology Quarterly* 50:186–204. [ADG]
Grimshaw, J. (1981) Form, function, and the language acquisition device. In: *The logical problem of language acquisition*, ed. C. L. Baker & J. J. McCarthy. MIT Press. [JG]
(1986) Subjacency and the S/S′ parameter. *Linguistic Inquiry* 17.364–69. [arDL, HWB, LR]
(1987) The components of learnability theory. In: *Modularity in knowledge representation and natural-language understanding*, ed. J. L. Garfield. MIT Press/Bradford Books. [EPS]
Gruber, J. (1967) Topicalization in child language. *Foundations of Language* 3:37–65. [GC]
Gueron, J. (1981) Logical operators, complete constituents, and extraction transformations. In: *Levels of syntactic representation*, ed. R. May & J Koster. Foris. [EPS]
Gumperz, J. J. & Hernandez-Ch., E. (1972) Bilingualism, bidialectalism, and classroom interaction. In: *Functions of language in the classroom*, ed. C B. Cazden, V. P. John & D. Hymes. Teachers College Press. [ADG]
Halliday, M. A. K. (1975) *Learning how to mean: Explorations in the development of language*. Edward Arnold. [ADG]
Hasher, L. & Zacks, R. T. (1984) Automatic processing of fundamental information. *American Psychologist* 39:1372–88. [IMS]
Hawkins, J. A. (1979) Implicational universals as predicators of word order change. *Language* 55:618–48. [aDL]
Hirsh-Pasek, K., Treiman, R. & Schneiderman, M (1984) Brown and Hanlon revisited: Mothers' sensitivity to ungrammatical forms. *Journal of Child Language* 11:81–88. [JG]
Hopper, P. & Traugott, E. C. (forthcoming) *Grammaticalization*. Cambridge University Press. [NV]
Horgan, D. (1978) The development of the full passive. *Journal of Child Language* 5:65–80. [CES]
Hornstein, N. & Lightfoot, D. W. (1987) Predication and PRO. *Language* 63:23–52. [aDL]
(1981) *Explanation in linguistics: The logical problem of language acquisition*. Longman. [aDL, RH]
Hornstein, N. & Weinberg, A. S. (1981) Case theory and preposition stranding. *Linguistic Inquiry* 12:55–91. [aDL]
Huang, J. (1982) Logical relations in Chinese and the theory of grammar. Unpublished doctoral dissertation, MIT. [aDL]
Hubel, S. H. & Wiesel, T. N. (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology* 160:106–54. [aDL]
Huttenlocher, J. (1974) The onset of language comprehension. In. *Theories in cognitive psychology*, ed. R. Solso. Erlbaum. [CES]
Hyams, N. (1983) The Pro-drop parameter in child grammars. *Proceedings of the West Coast Conference on Formal Linguistics*, Stanford. [arDL]
(1986) *Language acquisition and the theory of parameters*. Reidel. [GC]
(1987) The setting of the null subject parameter. A reanalysis. Paper presented at the Boston University Conference on Child Language Development, October. [JLM]
Hyman, L. & Comrie, B. (1981) Logophoric reference in Gokana. *Journal of African Languages and Linguistics* 3:19–37. [rDL, TW]
Jenkins, L. (1979) The genetics of language. *Linguistics and Philosophy* 3:105–19. [rDL]
Jerne, N. K. (1967) Antibodies and learning: Selection versus instruction. In: *The neurosciences: A study program*, ed. G. C. Quarton, T. Melnechuk & F. O. Schmitt. Rockefeller University. [aDL]
(1985) The generative grammar of the immune system. *Science* 229.1057–59. [aDL]
Jespersen, O. (1922) *Language: Its nature, development, and origin*. Allen & Unwin. [aDL, JDM]
Joshi, A. K. (1985) How much context sensitivity is needed for reasonable structural descriptions: Tree adjoining grammars. In: *Natural language parsing: Psycholinguistic, computational, and theoretical perspectives*, ed. D. Dowty, L. Karttunen & A. Zwicky. Cambridge University Press. [AKJ]
(1987) An introduction to tree adjoining grammars. In: *Mathematics of language*, ed. A. Manaster-Rammer. John Benjamins. [AKJ]
Joshi, A. K., Vijay Shanker, K. & Weir, D. (in press) Convergence of mildly context-sensitive grammar formalisms. In. *Natural language parsing*, ed. P. Sells & T. Wasow. MIT Press. [AKJ]
Kaplan, R. & Bresnan, J. (1982) Lexical-functional grammar: A formal system for grammatical representation. In: *The mental representation of grammatical relations*, ed. J. Bresnan. MIT Press. [WO]
Kayne, R. S. (1975) *French syntax: The transformational cycle*. MIT Press. [EPS]
(1981) ECP extensions. *Linguistic Inquiry* 12(1):93–133. Reprinted in R. S. Kayne (1983), *Connectedness and binary branching*. Foris. [aDL, EPS]
Keenan, E. & Comrie, B. (1977) Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8:63–100. [rDL, TW]
Klima, E. & Bellugi, U. (1966) Syntactic regularities in the speech of children In. *Psycholinguistic papers*, ed. J. Lyons & R. Wales Edinburgh University Press. [aDL]
Koch Christensen, K. (1982) On multiple filler-gap constructions in Norwegian. In. *Readings on unbounded dependencies in Scandinavian languages*, ed E. Engdahl & E Ejerhed. Almqvist & Wiksell. [HH]
Koopman, H. (1983) ECP effects in main clauses. *Linguistic Inquiry* 14:346–50 [aDL]
Koster, J. (1984) On binding and control. *Linguistic Inquiry* 15:417–59. [EW]
(1986) *Domain and Dynasties The radical autonomy of syntax*. Foris [EPS]
Kroch, A. S. (1987) Unbounded dependency and subjacency in a tree adjoining grammar In: *Mathematics of language*, ed. A. Manaster-Ramer. John Benjamins. [AKJ]
(1989) Function and grammar in the history of English periphrastic "do." In *Language variation and change*, ed. R. Fasold & D. Schiffrin. John Benjamins. [AK]
Kroch, A S. & Joshi, A. K. (in press) Linguistic relevance of tree adjoining grammars. In. *Linguistics and philosophy*. (Originally appeared as Technical Report MS-CIS-85-18, Department of Computer and Information Science, University of Pennsylvania, 1985.) [AKJ]
Kroch, A., Myhill, J. & Pintzuk, S. (1982) Understanding "do." *Proceedings of the 18th Annual Meeting of the Chicago Linguistic Society*. [AK]
Labov, W. (1972a) *Sociolinguistic patterns*. University of Pennsylvania Press. [AK]
(1972b) The internal evolution of linguistic rules. In: *Linguistic change and generative theory*, ed. R. P. Stockwell & R. K. S. Macaulay. Indiana University Press [ADG]
(1980) *Locating language in time and space*. Academic Press [AK]
(1986) Language structure and social structure. In: *Approaches to social theory*, ed. S. Lindenberg, J. S. Coleman & S Nowak. Russell Sage Foundation. [ADG]
Larson, R K. (1988) On the double object construction. *Linguistic Inquiry* 19:335–91 [YG]
Lasnik, H. (1981) Restricting the theory of transformations. A case study. In: *Explanation in linguistics: The logical problem of language acquisition*, ed. N. Hornstein & D. W. Lightfoot. Longman. [aDL]
(in press) On certain substitutes for negative data. In: *Essays on restrictiveness and learnability*, ed. H. Lasnik. Reidel [HL]
Lebeaux, D. S. (1987) Comments on Hyams. In. *Parameter setting*, ed. T Roeper & E. Williams. Reidel. [GC]
(1988) Language acquisition and the form of the grammar. Unpublished doctoral dissertation, University of Massachusetts, Amherst [rDL]
Lee, H. & Wexler, K (1987) The acquisition of reflexives and pronouns in Korean. From a cross-linguistic perspective Paper presented at the 12th annual Boston University Conference on Language Development. [rDL]
Levine, D. N. (1985) *The flight from ambiguity. Essays in social and cultural theory*. University of Chicago Press [ADG]
Lightfoot, D. W. (1979) *Principles of diachronic syntax*. Cambridge University Press. [aDL, NV]
(1982) *The language lottery. Toward a biology of grammars*. MIT Press. [arDL, WO]
(1986) Review of Culicover & Wilkins (1984). *Journal of Linguistics* 22:480–85. [aDL]
(1988) Creoles, triggers, and Universal Grammar. In: *A festschift for R. P. Stockwell*, ed. C. Duncan-Rose et al. Croom Helm. [aDL]
(in press) Obsolescence and Universal Grammar. In: *Proceedings of the fifth meeting of the International Conference on English Historical Linguistics*. Benjamins. [rDL]
(forthcoming) New parameter settings. [arDL]
Ludlow, P. & Neale, S. (1988) Indefinite descriptions. In defence of Russell, manuscript, Princeton University [SN]
MacWhinney, B. & Snow, C. E. (1985) The child language data exchange system. *Journal of Child Language* 12:271–96. [JLM]
Maling, J. (1986) Clause-bounded reflexives in modern Icelandic. In: *Topics in Scandinavian syntax*, ed. L. Hellan & K. Koch Christensen. Reidel. [HH]
Manzini, M.-R. & Wexler, K. (1987) Parameters, bindig theory, and learnability. *Linguistic Inquiry* 18:413–44. [RC, LR]
Maratsos, M. (1974) How preschool children understand missing complement sentences. *Child Development* 45.700–706. [CES]
Martinet, A. (1955) *Economie des changements phonétiques*. Francke. [NV]
Matthei, E. (1981) Children's interpretation of sentences containing reciprocals. In: *Language acquisition and linguistic theory*, ed. S. L. Tavakolian. MIT Press. [aDL]

Mazurkewich, I. & White, L. (1984) The acquisition of the dative alternation: Unlearning overgeneralizations. *Cognition* 16.261–83. [aDL]
McCarthy, J. & Hayes, P. (1969) Some philosophical problems from the standpoint of artificial intelligence. In. *Machine learning 4*, ed. B. Meltzer & D. Michie. Edinburgh University Press [RC]
McCawley, J. D. (1979) Language universals in linguistic argumentation. *Studies in the Linguistic Sciences* 8(2):205–19. (Reprinted in McCawley 1982b.) [JDM]
(1982a) The nonexistence of syntactic categories. In. *Thirty million theories of grammar*. University of Chicago Press. [JDM]
(1982b) *Thirty million theories of grammar*. University of Chicago Press. [JDM]
(1983) Towards plausibility in theories of language acquisition *Communication and Cognition* 16:169–83. [JDM]
(1985) Review of F. Heny and B Richards, *Linguistic categories. Language* 61:668–79. [JDM]
(1988) *The syntactic phenomena of English* University of Chicago Press. [JDM]
McNeill, D. (1966) Developmental linguistics. In *The genesis of language: A psycholinguistic approach*, ed F. Smith & G. A. Miller MIT Press [aDL, CES]
Mehler, J. (1974) Connaître par desapprentissage. In: *L'unité de l'homme*, ed. E. Morin & M. Piatelli-Pamarini. Editions du Seuil. [aDL]
Morgan, J L. (1986) *From simple input to complex grammars*. MIT Press. [aDL, JLM, EPS]
Nelson, K. E. (1982) Theories of the child's acquisition of syntax. *Annals of the New York Academy of Sciences* 345:45–69. [CES]
(1987) Reflections on a rare-event cognitive comparison theory of language acquisition. In: *Children's language*, vol. 6, ed. K. E. Nelson & A. van Kleeck. Erlbaum. [CES]
Newport, E. C. (1977) Motherese· The speech of mothers to young children. In: *Cognitive theory*, vol. 2, ed. N Castellan, D Pisoni, & G. Potts. Erlbaum. [EPS]
Newport, E. C., Gleitman, H. & Gleitman, L. (1977) "Mother, I'd rather do it myself": Some effects and non-effects of maternal speech style. In: *Talking to children. Language input and acquisition*, ed C. E. Snow & C. A. Ferguson. Cambridge University Press [aDL, RC]
Noble, S (1985) To have and have got. Paper delivered at the New Ways of Analyzing Variation in English (NWAVE) XIV Conference at Georgetown University. [AK]
Obenauer, H.-G. (1984) On the identification of empty categories. *Linguistic Review* 4:153–202. [EPS]
Ochs, E. (1986) Introduction. *Language socialization across cultures*, ed. B. B. Schieffelin & E. Ochs. Cambridge University Press [ADG]
O'Grady, W. (1987) *Principles of grammar and learning*. University of Chicago Press. [WO]
(forthcoming) Categories and case: The sentence structure of Korean. [WO]
Oliveira e Silva, G. (1982) Estudo da Regularidade na Variação dos Possessivos no Português do Rio de Janeiro. Tese du Doutorado em Letras. Universidade Federal do Rio de Janeiro. [AK]
Osherson, D., Stob, M. & Weinstein, S. (1986) *Systems that learn*. MIT Press. [JLM]
Osherson, D. & Wasow, T. (1976) Task-specificity and species-specificity in the study of language. *Cognition* 4:203–14. [WO]
Penner, S. (1987) Parental responses to grammatical and ungrammatical child utterances. *Child Development* 58.376–84 [JG]
Pesetzky, D. (1982) Paths and categories. Unpublished dissertation, Massachusetts Institute of Technology. [HH]
Phinney, M. (1987) The pro-drop parameter in second language acquisition. In: *Parameter setting*, ed. T. Roeper & E. Williams. Reidel [HWB]
Piattelli-Palmarini, M. (1986) The rise of selective theories A case study and some lessons from immunology. In. *Language learning and concept acquisition*, ed. W Demopoulos & A. Marras. Ablex [aDL]
Pica, P. (1987) On the nature of the reflexivisation cycle. *Proceedings of the North Eastern Linguistic Society*, 17. [rDL, LR]
Pinker, S. (1979) Formal models of language learning. *Cognition* 7:217–83. [JG]
(1984) *Language learnability and language development*. Harvard University Press. [JG, HH, CES]
(1988) Resolving a learnability paradox in the acquisition of the verb lexicon. In. *Lexicon project working papers 17*. Center for Cognitive Science, MIT. [YG]
(1989) *Learnability and cognition: The acquisition of argument structure* MIT Press. [JG]
Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1972) *A grammar of contemporary English*. Longman. [IMS]
Radford, A (1988) Small children's small clauses. *Transactions of the Philological Society* 86.1–43. [rDL, GC]
Randall, J. (1985) Retreat routes. Paper presented at the Boston University Conference on Language Development, October. [arDL, YG]
Reinhart, T. (1975) A note on the two-comp hypothesis. Unpublished paper. Massachusetts Institute of Technology [JK]
Rizzi, L. (1982a) *Issues in Italian syntax* Foris [arDL, RF, LR]
(1982b) Comments on Chomsky's "On the representation of form and function." In. *Perspectives on mental representation*, ed J Mehler, E C. T Walker & M. Garrett Erlbaum. [aDL, GC, RH]
(1986) Null objects in Italian and the theory of pro. *Linguistic Inquiry* 17:501–57. [GC, LR]
Roberts, I (1985) Agreement parameters and the development of English modal auxiliaries. *Natural Language and Linguistic Theory* 3.21–58. [AK]
Roeper, T. (1979) Children's syntax. Mimeo, University of Massachusetts, Amherst. [aDL]
Roeper, T & Williams, E., eds. (1987) *Parameter setting*. Reidel. [EW]
Romaine, S. (1985) Why the problem of language acquisition should not be explained logically *Studies in Language* 9(2):255–70. [RH]
Ross, J. R. (1967) Constraints on variables in syntax. Doctoral dissertation, Massachusetts Institute of Technology. [WW]
(1973) The penthouse principle and the order of constituents. In: *You take the high node and I'll take the low node: Papers from the Comparative Syntax Festival*, ed. C Corum, T. C. Smith-Stark & A. Weiser Chicago Linguistics Society. [aDL]
Rouveret, A. & Vergnaud, J.-R. (1980) Specifying reference to the subject: French causatives and conditions on representations. *Linguistic Inquiry*. 11 97–202 [aDL]
Russell, B. (1919) *Introduction to mathematical philosophy*. Allen & Unwin. [SN]
Sachs, J. & Truswell, L. (1978) Comprehension of two-word instructions by children in the one-word stage. *Journal of Child Language* 5:17–23. [CES]
Safir, K J. (1985) *Syntactic chains*. Cambridge University Press. [EPS]
Sainsbury, R. (1971) The "Feature Positive Effect" and simultaneous discrimination learning. *Journal of Experimental Child Psychology* 11·347–56. [aDL]
(1973) Discrimination learning utilizing positive or negative cues. *Canadian Journal of Psychology* 27 46–57. [aDL]
Sankoff, G. (1980) *The social life of language* University of Pennsylvania Press. [ADG]
Schauber, E (1979) *The syntax and semantics of questions in Navajo*. Garland. [WW]
Schieffelin, B. B. & E Ochs, eds. (1986) *Language socialization across cultures* Cambridge University Press. [ADG]
Schlesinger, I. M. (1977) *Production and comprehension of utterances*. Erlbaum. [IMS]
(1982) *Steps to language: Toward a theory of language acquisition*. Erlbaum. [IMS]
(1985) *Instruments as agents*. Working Paper No. 15, The Goldie Rotman Center for Cognitive Science in Education. Hebrew University, Jerusalem. [IMS]
(1988) The origin of relational categories. In: *Categories and processes in language acquisition theory*, ed. Y. Levy, I M. Schlesinger & M. D. S. Braine. Erlbaum. [IMS]
Sells, P. (1987) Aspects of logophoricity. *Linguistic Inquiry* 18:445–79. [rDL, TW]
Sigurjonsdottir, S. & Hyams, N (1988) The acquisition of reflexives and pronouns by Icelandic children. *Papers and Reports on Child Language Development*, 27 Stanford University [rDL]
Slobin, D. (1968) Recall of full and truncated passives in connected discourse *Journal of Verbal Learning and Verbal Behavior* 7:876–81. [CES]
Slobin, D I. (1985a) Crosslinguistic evidence for the language-making capacity. *The crosslinguistic study of language acquisition*. Vol. 2: *Theoretical issues*, ed. D. I. Slobin. Erlbaum. [ADG]
(1985b) *The crosslinguistic study of language acquisition*. Vol. 1: *The data*. Erlbaum. [ADG]
Smith, F. & Miller, G. (1966) *The genesis of language. A psycholinguistic approach*. MIT Press. [MPL]
Snow, C. (1972) Mothers' speech to children learning language. *Child Development* 43:549–65. [CES]
(1977) Mothers' speech research. From input to interaction. In: *Talking to children. Language input and acquisition*, ed. C. E. Snow & C. A. Ferguson. Cambridge University Press. [aDL, ADG, CES]
(1982) The uses of imitation. *Journal of Child Language* 8:205–12. [CES]
(1983) Saying it again: The role of expanded and deferred imitations in

language acquisition. In: *Children's language*, vol. 4, ed. K. E Nelson Gardner Press. [CES]

(in press) Understanding social interaction and language development: Sentences are not enough. In. *Interaction in human development*, ed. M Borstein & J. Bruner. Erlbaum. [CES]

Snow, C E. & C. A. Ferguson, eds. (1977) *Talking to children language input and acquisition*. Cambridge University Press [ADG]

Snow, C., Perlmann, R. & Nathan, D. (1987) Why routines are different· Towards a multi-factors model of the relation between input and language acquisition. In: *Children's language*, vol. 6, ed. K. E. Nelson & A van Kleeck. Erlbaum. [CES]

Sportiche, D. (1981) On bounding nodes in French. *Linguistic Review* 1.219–46. [aDL, RF]

Stabler, E. P., Jr. (forthcoming) The logical approach to syntax. [EPS]

Tomasello, M. (1987) Learning to use prepositions. A case study *Journal of Child Language* 14:79–98. [CES]

(forthcoming) *Verbs and grammar in early child language. A case study*. Cambridge University Press. [CES]

Torrego, E. (1984) On inversion in Spanish and some of its effects. *Linguistic Inquiry* 15:103–29. [HWB, WW]

Travis, L. (1987) Parameters of phrase structure. Mimeo, McGill University. [aDL]

Valiant, L. G. (1984) A theory of the learnable. *Communications of the Association for Computing Machinery* 27:1134–42. [AKJ]

Van Oosten, J. (1977) Subjecthood and agenthood in English *Papers from the thirteenth regional meeting, Chicago Linguistic Society*. University of Chicago Press. [IMS]

Vincent, N. (1987) English seem and Italian sembrare. The origin of thematically empty subjects. Paper delivered at the Fifth International Conference of English Historical Linguistics,' Cambridge, April. [NV]

Warner, A. (1983) Review article on Lightfoot (1979). *Journal of Linguistics* 19.187–209. [aDL]

Weinberg, A. (1987) Comments on Borer and Wexler In. *Parameter setting*, ed T Roeper & E. Williams. Reidel. [rDL, GC]

Weinberg, A S., Aoun, J., Hornstein, N & Lightfoot, D. W. (1987) Two types of locality. *Linguistic Inquiry* 18:537–77 [aDL]

Weinreich, U., Labov, W & Herzog, M. I (1968) Empirical foundations for a theory of language change. In: *Directions for historical linguistics*, ed. W P. Lehmann & Y. Malkiel. University of Texas Press. [ADG]

Wells, G. (1981) *Learning through interaction. The study of language development*. Cambridge University Press. [arDL, CES]

Wexler, K. & Culicover, P. W. (1980) *Formal principles of language acquisition*. MIT Press. [arDL, RC, JG, HL, JLM, WW]

Wexler, K. & Manzini, R (1987) Parameters and learnability in the binding theory. In. *Parameter setting*, ed. T Roeper & E. Williams Reidel. [RC]

White, L. (1987) A note on Phinney. In. *Parameter setting*, ed. T. Roeper & E. Williams Reidel. [HWB]

Wilcoxon, S. M. (1988) Binding in Mandarin Chinese. Another look at *ziyi*. Unpublished manuscript. Department of Linguistics, University of Texas at Austin. [rDL, CLB]

Williams, E. (1987) Introduction. In: *Parameter setting*, ed. T Roeper & E. Williams. Reidel. [HWB]

Williams, M. M. (1973) A case of unmarked subordination in Tuscarora. In: *You take the high node and I'll take the low node. Papers from the Comparative Syntax Festival*, ed. C Corum, T C. Smith Stark & A. Weiser. Chicago Linguistics Society, University of Chicago. [TW]

# Please send BBS your email address

All BBS Associates and any non-Associates who have served as referees, commentators or authors, or who are qualified and interested in serving as referees or commentators for BBS:

Please send us your electronic mail address, if you have one. (If you don't have one, you are strongly urged to look into the advantages of getting one -- not only for BBS's sake!)

BBS is implementing more and more of its peer communication functions by electronic mail. This not only increases the speed and efficiency of BBS's interaction with the world biobehavioral and cognitive science community, but it dramatically increases its scope and range as well. Abstracts can be circulated by email in advance to allow potential commentators to nominate themselves. Referee reports can be submitted by email. The BBS Associateship can be more representatively canvassed to determine what topics and authors they would like to see treated in BBS. New Associates can be nominated by email, etc.

Electronic mail addresses can be sent to our regular mail address. (Any available *departmental* or *institutional email directories* would be very helpful too.)

**Behavioral and Brain Sciences**
**20 Nassau Street, Room 240**
**Princeton NJ 08542**

Or they can be sent by electronic mail to any of the following electronic mail addresses. (Because email is not yet reliable, please try several until you receive confirmation that your message has been received.)

**INTERNET**

**harnad@confidence.princeton.edu**
**harnad@princeton.edu**
**harnad@elbereth.rutgers.edu**
**srh@flash.bellcore.com**

**BITNET**

**harnad@pucc.bitnet**

**UUCP**

**harnad@princeton.uucp**
**princeton!confidence!harnad**

**CSNET**

**harnad%confidence.princeton.edu@relay.cs.net**

Along with your email address you are encouraged to include you suggestions about current BBS editorial policy and directions you would like to see BBS take in the future.